

Yale University

EliScholar – A Digital Platform for Scholarly Publishing at Yale

Yale Graduate School of Arts and Sciences Dissertations

Spring 2021

Living and Judging: Studies in Pragmatic Reasoning in Philosophy, Literature, and Law

Tobias Kuehne

Yale University Graduate School of Arts and Sciences, tobi.kuehne@gmail.com

Follow this and additional works at: https://elischolar.library.yale.edu/gsas_dissertations

Recommended Citation

Kuehne, Tobias, "Living and Judging: Studies in Pragmatic Reasoning in Philosophy, Literature, and Law" (2021). *Yale Graduate School of Arts and Sciences Dissertations*. 70.
https://elischolar.library.yale.edu/gsas_dissertations/70

This Dissertation is brought to you for free and open access by EliScholar – A Digital Platform for Scholarly Publishing at Yale. It has been accepted for inclusion in Yale Graduate School of Arts and Sciences Dissertations by an authorized administrator of EliScholar – A Digital Platform for Scholarly Publishing at Yale. For more information, please contact elischolar@yale.edu.

Abstract

Living and Judging: Studies in Pragmatic Reasoning in Philosophy, Literature, and Law

Tobias Kuehne

2021

This dissertation presents five studies on the problem of formalization. Every body of knowledge—whether it be scientific, philosophical, literary, practical, or personal—has its own conventions by which that knowledge is generated, arranged, reworked, presented, and defended. But formalization is rarely without its blind spots or glitches. Sometimes, a body of knowledge seeks to mask its techniques of formalization to cover up its own contingency or lack of ultimate answers. In this process, rhetoric plays an elusive dual role. Rhetoric sometimes contributes to formalization, while sometimes undermining it. Studying the rhetoric of a body of knowledge thus serves as an access point to its problems of formalization.

The case studies in this dissertation approach the problem of formalization from different angles and in different fields. Chapters one and two examine the work of Friedrich Nietzsche and Edmund Husserl, two philosophers who exposed the problem of formalization in philosophy and science, and grappled with its implications for lived experience. For Nietzsche, philosophical dialectical reasoning is merely one rhetorical technique among many. By rhetorically performing dialectical reasoning in the *Genealogy of Morality*, Nietzsche lays bare how the ascetic ideal that animates dialectics runs into an aporia: When its negatory power turns back against itself, the negation must fail either because it cannot negate itself, or because it persists in *performing* a successful self-negation. Uncovering aporias in formalized systems of knowledge is not only a central impetus to Nietzsche's thought, but also paves the way to an

ethical life. Only without the guardrails of pre-formalized imperatives can the freely philosophizing spirit devise an ethics that is truly its own and deserving of the name.

Chapter two turns to Edmund Husserl's theory of the lifeworld. In *The Crisis of the European Sciences*, Husserl develops the lifeworld to elucidate how human experience became untethered from scientific thought. For Husserl, science after Galileo took on an attitude that treats the world as a vast technical structure that can be precisely measured, understood, and formalized in a calculus that proceeds from general principles to theorems. Human experience, however, proceeds from the particular to the general. The *modus operandi* of human experience—and the lifeworld—is thus thoroughly inductive. This inductive operativity is an intentional, historical practice. Always anticipating that the incomplete experience of the moment will be expanded, completed, or corrected by future experience, human subjectivity is both inherently historical and scientific.

Chapters three and four discuss Franz Kafka and Jacques Lacan, two (broadly) literary authors who push the problem of formalizing the intersubjective relationship to its breaking point and puts its implosion to creative use. Kafka's *Letter to the Father* defines a deceptive logical calculus that does the opposite of what it purports to do. Modeling this logical structure with the method of gödel numbering, this chapter traces how this logic deliberately defeats itself. In purporting to justify himself, Kafka weaves reproaches into the letter to tempt his father to justify *himself*. In laying this trap, Kafka tries to trick his father into adopting the rhetorical register of guilt in which Kafka is stuck. The letter is thus an attempt by Kafka to re-formalize their relationship. The attempt itself—and its failure—propels and sustains Kafka's writing.

Chapter four dives into Jacques Lacan's early attempts to formalize psychoanalysis into a science. In his early seminars, Lacan lays out how theories of intersubjectivity run into an

infinite regress if they only conceptualize small “others” who attempt to divine what each of them thinks. To break this infinite regress, Lacan theorizes a symbolic order of play that always already structures intersubjective relations. This formalization, however, is by no means a predictable, predetermined process. Signifiers without a signified disrupt the symbolic chain and impel it into unforeseen directions. In the absence of a signified, the speaking subject produces signifiers to fill the void. As all of them fail to take the position of the signified (none of them can be taken literally), the speaking continues.

Chapter five presents a piece of legal history that traces the genesis of qualified immunity doctrine in the United States. While the doctrine’s initial purpose was to preserve judicial flexibility in cases where private individuals sued government officials, qualified immunity soon took a path of formalizing itself into a rigid standard that effectively shields almost all illegal conduct by government officers. Qualified immunity thus tells a story of how problematic formalization can become in the law—and how dangerous it can be to ignore.

Living and Judging: Studies in Pragmatic Reasoning in Philosophy, Literature, and Law

A Dissertation
Presented to the Faculty of the Graduate School
of
Yale University
in Candidacy for the Degree of
Doctor of Philosophy

by
Tobias Kuehne

Dissertation Director: Rüdiger Campe

June 2021

© 2021 by Tobias Kuehne
All rights reserved.

TABLE OF CONTENTS

INTRODUCTION: RHETORIC AND THE PROBLEM OF FORMALIZATION.....	10
I. WHAT RHETORIC IS NOT	12
II. CLASSICAL DEFINITIONS AND TYPOLOGIES	14
1. The Sophists and the Art of Speaking: Forensic, Deliberative, and Epideictic Rhetoric	14
2. Plato: Episteme Over Doxa	15
3. Aristotle: Finding the Available Means of Persuasion in a Given Situation	16
4. Cicero: Institutionalized Adversarial Testing.....	19
5. Quintilian: The Science of Speaking Well	20
III. RHETORIC AND THE PROBLEM OF FORMALIZATION: THREE DIMENSIONS	21
1. Rhetoric and Knowledge: Means of Figuration	21
2. Rhetoric and Ethics: The Silent Call to Responsibility	23
3. Rhetoric and History: Activating the Potential of the Moment	28
IV. THE CHAPTERS OF THIS DISSERTATION.....	30
WORKS CITED.....	35
 CHAPTER 1: NIETZSCHE: RHETORIC AGAINST A FORMALIST WAY OF LIFE	38
PART I: THE RHETORICAL PERFORMANCE OF PHILOSOPHY	39
1. The Offense of Hegelianism	39
2. Marx: The Chiasmic Structure of Dialectics	43
3. Nietzsche and the Historicity of Dialectics.....	48
4. Vigor and Sickness: Thinking in a New Key	51
5. Gesturing into (the) Nothing: Aposiopesis and the Rhetoric of Silence.....	60
PART II: LIVING IN THE ABSENCE OF SYSTEMS: NIETZSCHE’S ETHICS OF DANGER.....	63
1. Introduction.....	63
2. Establishing the Task of Confronting Danger: Nietzsche’s Early Works	66
3. Embarking on Perilous Seas: The Demise of Metaphysics and the Ethical Question	76
4. Concluding Remarks: Danger’s Strategic Role in Nietzsche’s Philosophy.....	83
WORKS CITED.....	86
 CHAPTER 2: THE INDUCTIVITY OF THE LIFEWORLD IN HUSSERL’S LATE THOUGHT .	91
INTRODUCTION	91
I. MATHEMATIZATION AND INDUCTION’S TENDENCY TO TRANSFORM INTO DEDUCTION	95
1. Umstellung der Einstellung: Three Moments.....	95
2. Vignette: The Genesis of Mathematical Induction as an Exemplar for the Attitudinal Shift	100
II. THE INDUCTIVITY OF THE LIFEWORLD—AND OF HUSSERL’S THINKING	103
1. The Lifeworld in Section 9: Anticipation, Horizon, Uniform Causal Style	103
2. The Humean Problem and the Reentry of the Lifeworld	107
3. The Motivational Problem.....	110
III. HUSSERL AND HUME’S PROBLEM OF INDUCTION.....	113
1. The Humean Account.....	113
2. Husserl’s Response to Hume’s Problem of Induction.....	116

IV.	THE MISSING LINK: ARISTOTELIAN NOUS AND EPAGOGÉ	119
1.	The Role of <i>Nous</i> in Grasping the Universal from Repeated Sense Perception	119
2.	Epagoge, the Analogy of the Rout, and Exemplarity.....	126
3.	Example and Exemplar.....	128
V.	EXEMPLARITY, INDUCTION IN HUSSERL’S THEORY OF LIFE AND INDUCTION: PERFORMING THE EPOCHÉ	133
	WORKS CITED.....	143
 CHAPTER 3: FORMALIZATION’S STRATEGIC PARADOXES IN KAFKA’S DISCOURSE OF GUILT		147
I.	THE VEXATIOUS QUESTION OF GENRE	147
II.	APPLYING GÖDEL’S CALCULUS TO KAFKA’S LOGIC	151
1.	Justifying the Use of Formal Language	151
2.	Formalizing the Father’s and Kafka’s Language.....	153
3.	Conditions for the Diagonal Lemma to be Derivable	156
4.	Applying the Diagonal Lemma to Kafka	159
5.	Interpreting the Result of the Diagonal Lemma.....	161
III.	NACH LOGIK DIE SINTFLUT: THE LOGIC OF VORWURF, RESPONSE, AND GUILT	163
1.	Eliciting a Response.....	163
2.	Kafka’s Lawyerly Ruses.....	165
3.	Dropping the Pretense	168
IV.	COMING TO REST?	172
	WORKS CITED.....	176
 CHAPTER 4: LACAN, FORMALIZATION, AND PSYCHOANALYSIS		178
	LACAN AND THE PROBLEM OF FORMALIZATION.....	178
	PART I: LACAN, GAME THEORY, AND THE PROJECT OF FORMALIZATION.....	180
1.	Monkey See, Monkey Do	180
2.	The Intersubjective Prison—and Its Dilemmas	182
3.	“To the Exit!” — Time, Exigency, and a Glimpse of the Other	190
4.	Beyond Intersubjectivity: The Other, the Symbolic, Play.....	195
	PART II: LACAN’S SHIFTING APPROACH TO FORMALIZATION	198
1.	Arresting the Attention	199
2.	Algorithmicity and a Blind Alley.....	201
3.	Reconstructing the Stakes of Lacan’s Mathematics and Science.....	204
4.	Mathemes Avant la Lettre and a Propulsive Effect.....	211
	WORKS CITED.....	217
 CHAPTER 5: FORGETTING MARBURY’S LESSON: A HISTORY OF QUALIFIED IMMUNITY		221
	INTRODUCTION	222
1.	The Current State of Qualified Immunity Doctrine.....	222
2.	The Current State of Qualified Immunity Scholarship.....	225
3.	The Contribution: An Institutional History and the <i>Marburian</i> Move.....	228
I.	<i>PIERSON V. RAY</i> AND THE MARBURIAN MOVE.....	231

1.	<i>Pierson's</i> Oddly Conservative Holding.....	231
2.	<i>Pierson's</i> Historical Context.....	234
3.	<i>Pierson's</i> Marburian Move.....	239
II.	FOLLOWING <i>MARBURY</i> (MOSTLY): CRAFTING A NEW STANDARD FROM <i>SCHEUER V. RHODES</i> TO <i>HARLOW V. FITZGERALD</i>	241
1.	<i>Scheuer v. Rhodes</i> : Broadening the Defense and Raising New Questions.....	241
2.	Towards an Objective Standard: The “Promise of § 1983”.....	246
3.	Towards a Uniform Standard: “My Vote or Byron’s!”.....	256
4.	Summing Up: Qualified Immunity’s <i>Marburian</i> Origins.....	266
III.	LOSING SIGHT OF <i>MARBURY</i> : THE DOOMED PROJECT OF <i>SAUCIER V. KATZ</i>	269
1.	<i>Marburian</i> Moves in Peripheral Questions.....	270
2.	Two Departures from <i>Marbury</i>	274
3.	Dead on Arrival: <i>Saucier’s</i> Internal Tensions.....	282
IV.	<i>MARBURY</i> TO THE WINDS: <i>PEARSON V. CALLAHAN</i> AND BEYOND.....	290
1.	<i>Pearson’s</i> Retreat.....	290
2.	<i>Pearson’s</i> Dodge.....	291
3.	<i>Pearson’s</i> Aftermath.....	293
V.	UPSHOTS: DECIDING TO DECIDE.....	296

LIST OF ILLUSTRATIONS

Figure 1: “Odd or Even” Payoff Matrix.	177
Figure 2: Prisoner’s Dilemma Payoff Matrix.	180
Figure 3: Lacan’s Saussurean algorithm in <i>IL</i>	195
Figure 4: Lacan’s development upon the algorithm.	195
Figure 5: The metonymy function.	195
Figure 6: The metaphor function.	195

ACKNOWLEDGMENTS

This dissertation would not have been possible without the generous support of the many faculty and advisers who guided me throughout the years. My thanks go to Jan Hagens, Charles Hill, Carol Jacobs, Nancy Levine, Rainer Nägele, Paul North, Martin Ruehl, Henry Sussman, Katrin Trüstedt, Kirk Wetters, and John Witt for their feedback, encouragement, and patience. I owe special thanks to Rüdiger Campe, who has been a nurturing reader since my undergraduate time at Yale.

Family members who have been steady sources of support were my mother, Simone Kühne; my father, Matthias Kühne; and my stepmother, Margit Salatzki. Friends who inspired, encouraged, and guided me over the years were Natalia Emanuel, Hannah Kleinfeld, Ela Leshem, Lynette Lim, Ryan Martins, Bennett Ostdiek, Thiti Owlarn, Sean Park, Jonas Rosenbrück, William Stone, Jan Claas van Treeck, and Nick Venable. And of course, my most special thanks go to Rachel Sommers.

INTRODUCTION: RHETORIC AND THE PROBLEM OF FORMALIZATION

The study of rhetoric is an amorphous field. George Kennedy has pointed out that “[t]he term *rhetoric* has clearly had different meanings in different historical cultures, and the phenomena we call ‘rhetoric’ have been called different things at different times” (Kennedy 1-2). Not only has the meaning of rhetoric shifted over the millennia, but rhetoric has often been polemicized (Goodrich 614) and shunted to the margins (Hariman 38-39). Against these difficulties, it seems nearly impossible to present a definition of rhetoric that is both precise and uncontroversial.

Yet, as Robert Scott has remarked, “[t]he problem of defining ‘rhetoric’ seems irresistible to rhetoricians” (Scott 1973, 81). Recognizing the definitional difficulty, Kennedy has proposed to “define a ‘genus’ of which the various historical meanings of rhetoric are ‘species,’” whose common denominator is “the energy inherent in communication” (Kennedy 1-2). Scott similarly responds to the definitional problem by proposing an understanding of rhetoric in three broad senses: its timeless argumentative forms, the social values it embodies, and the everyday way of life it engenders (Scott 1973, 93-94). Mark Porrovecchio and Celeste Condit, observing that historical definitions of rhetoric have ranged from narrowly encompassing style and delivery to broadly encompassing all types of discourse, suggest to abandon the question of what rhetoric *is* and instead “ask a more inclusive and proactive question: What can a ‘rhetoric’ be?” (Porrovecchio & Condit 15). The persistent attempt—and failure—to define rhetoric point to a deep uncertainty over rhetoric’s task, status, and legitimacy.

The attempt to legitimize rhetoric by way of an authoritative definition is prone to failure because rhetoric finds itself in a dichotomy with a term that is always already superior to it. That superior term has variously appeared as “knowledge,” “science,” or “philosophy.”

“[R]hetoric has been placed on the defensive” throughout modern history because of its dubious epistemic value (Goodrich 613). Indeed, rhetoric’s deficiency in conveying truth has caused “embarrassment” for its practitioners ever since it originated among the sophists in Magna Graecia (Scott 1967, 9). Anthony Kronman’s attempt to define the proper field of rhetoric as political truths in which scientific certainty cannot be achieved but raw personal preferences should not run amok is affirms the rhetoric-knowledge dichotomy and tries to salvage rhetoric from that defensive position (Kronman 683-87). Brian Vickers, too, presents his seminal *In Defence of Rhetoric* as an attempt to “remove the misapprehensions and prejudices that still affect our appreciation of rhetoric” (Vickers 1).

This dissertation takes a different approach. Instead of understanding rhetoric as a dichotomous (and inferior) term to knowledge, it treats rhetoric as an entry point to a problem that *all* knowledge has to contend with: the problem of formalization. Approaching rhetoric as an object of study that requires clarification on “what it is” and differentiation from “what it is not” replicates the metaphysical distinctions between science and art, episteme and techne, substance and accidents. To continue operating under this set of dichotomies thus reproduces the same baggage with which Nietzsche, Heidegger, and their successors sought to dispense. Instead, this dissertation suggests that rhetoric poses, presents, and performs the problems that any field of knowledge undergoes in formalizing itself. Rhetoric is not one term in a dyad, but a zone of instability in which a field of knowledge is constantly questioned, negotiated, arranged, and rearranged. Rhetoric is thus not so much an object or technique in need of definition, but an access point to how a field of knowledge contends with the problem of formalizing itself.

The chapters that follow will provide case studies that approach the problem from different angles and in different fields. Chapters one and two study the work of Friedrich

Nietzsche and Edmund Husserl, two philosophers who exposed and grappled with the problem of formalization. Chapters three and four discuss Franz Kafka and Jacques Lacan, two (broadly) literary authors who push the problem of formalizing the intersubjective relationship to its breaking point and harness its implosion for generative purposes. Chapter five takes yet another approach. As a piece of legal history, it traces the genesis of qualified immunity doctrine in the United States. While the doctrine's initial purpose was to preserve judicial flexibility in cases where private individuals sued government officials, qualified immunity soon took a path of formalizing itself into a rigid standard that effectively shields almost all illegal conduct by government officers. Qualified immunity thus tells a story of how problematic formalization can become in the law—and how dangerous it can be to ignore.

But before embarking on those studies, this introduction will work out how, throughout its history, rhetoric has always been an indicator and symptom of the problem of formalization. To this end, the introduction will 1) briefly sketch rhetoric's ancient and recent history; 2) derive the perennial instabilities that rhetoric exposes and creates in knowledge, ethics, and history; and 3) clarify my understanding of rhetoric as a symptom of the problem of formalization, not a dichotomous counter term to knowledge, philosophy, or science.

I. WHAT RHETORIC IS NOT

Rhetoric is often negatively defined as a type of discourse that differs from some other type of discourse. Rhetoric is *not* science; it is *not* philosophy (or dialectic); and it is *not* poetry (Porrovecchio 155). This is to say that rhetoric is not a stable, systematically structured body of truths. Science, philosophy, and poetry purport to contain true and certain knowledge (albeit each of their own kind), while rhetoric is the realm of the probable and of opinion (Barilli 6-13). In rhetoric, there is no knowledge other than what rhetor and audience agree on through

persuasion, argumentation, consensus, demagoguery, or whichever technique the rhetor happens to employ (Parker 70, Remer *passim*, Conley 299-304). By contrast, the truths of science, philosophy, and poetry purport not to depend on an accepting audience. Galileo's proof of heliocentrism holds true regardless of Pope Urban VIII's personal views or Galileo's recantation. A philosophical argument is true or false, independently of how persuasive it may be to a listener. And even a poem's truth is not tied to whether a reader recognizes or understands it (Porrovecchio & Condit 15-16). Non-rhetorical disciplines lay their own (often universal, sometimes totalizing) claims to truth. While they are adaptable over time, they have audience-independent principles of stability, timelessness, objectivity, and organization. In other words, non-rhetorical disciplines aspire to formalization.

Rhetoric, by contrast, is a linguistic practice, a set of techniques to be put to use. Rhetoric is a type of discourse that *emerges* in a back-and-forth between an audience and a rhetor who tries to have an effect on that audience—changing minds, evoking emotions, or inducing actions. This raises interesting questions about rhetoric's relationship with politics, ethics, culture, language, and legitimacy, some of which will be addressed below. For now, I take rhetoric's fundamental distinction from philosophy, science, and poetry as a starting point to look at a few canonical accounts of rhetoric by the sophists, Plato, Aristotle, Cicero, and Quintilian. Each of them articulated a distinctive, influential position in the debate on rhetoric that continue to inform and reappear in modern discussions. These accounts are not meant to be in-depth reconstructions, but to serve as a springboard for identifying the problems that rhetoric raises for formalization.

II. CLASSICAL DEFINITIONS AND TYPOLOGIES

1. The Sophists and the Art of Speaking: Forensic, Deliberative, and Epideictic Rhetoric

Rhetoric did not emerge as a theoretical discipline of systematized knowledge, but as a practice adapted to specific needs. One such need was to structure the adjudication of legal disputes in a popularly constituted forum. After the democratic revolts in the city states of Magna Graecia in the sixth century BCE, legal disputes were decided by juries of lay citizens or judges chosen by lot (Gagarin 2017, 46-47). As professionalized legal representation was unknown at the time, laypeople had to persuade an audience of their peers that their version of the events was more plausible than the other side's, and that their legal claims were valid (Gagarin 2017, 44). Wealthy individuals thus turned to the sophists for instruction in composing, memorizing, and delivering speeches, as well as public debate (Gagarin 2017, 47). Citizens' need of having to persuade lay decisionmakers in a legal forum thus gave rise to the genre of forensic (or judicial) rhetoric. Rhetoric's task in this adversarial legal setting was, to speak with Protagoras, "make the weaker argument appear stronger" (Gagarin 2002, 24-26).¹

The second genre of rhetoric that emerged in the Greek city states was deliberative (or political) rhetoric. Orators spoke in the public forum to persuade an audience to make a decision or take action on a question of policy (Hollander 179). Persuasion aimed to induce effects and action. The third genre was epideictic rhetoric (or praise). The speaker had to harness his skills to highlight the virtues of a person, a group, or an object (ibid.).

¹ Critics of rhetoric have taken this formulation to mean that rhetoric is manipulative: to make the weaker of two arguments appear to be the stronger one. Others have replied that Protagoras merely meant that rhetoric strives to present any given argument in its strongest possible light (Gagarin 2002, 24-26).

The three genres fulfilled important functions in a heavily litigious society. Forensic rhetoric was the engine for a developing body of law; deliberative rhetoric was the catalyst for coordinated action; and epideictic rhetoric built and affirmed community values. To be sure, the relationship between the three genres is more complex than this typology suggests. For example, “[t]he epideictic form is eccentric vis-à-vis the triadic whole” insofar that epideictic rhetoric seeks to persuade by seduction, while the other two persuade by agonistic refutation (Campe 2015, 45). At the same time, forensic and deliberative rhetorics can be inflected by epideictic traits (ibid.). Each type of rhetoric constituted a linguistic art that generated practices and communal knowledge by vying for audience approval. At the same time, rhetoric was a domain of contestation and competition. Rhetoric’s emergent, audience-dependent, and competitive situation thus made it an agent of both stability and instability.

As much as a single definition is possible, one might therefore adopt John Poulakos’s suggestion that sophistic “[r]hetoric is the art which seeks to capture in opportune moments that which is appropriate and attempts to suggest that which is possible” (Poulakos 36). Resisting the view that sophistic rhetoric consisted of a set of rigid, formalistic rules, Poulakos emphasizes how much their practice promoted creativity and personal expression, required a skill of saying the right thing at the right moment, and had the potential to shift the audience’s point of view by opening their minds to new possibilities (ibid. 36-42). From the beginning, rhetoric as a practical, professionalized art was both an agent of formalization and in tension with it.

2. Plato: Episteme Over Doxa

From its inception, rhetoric and the sophists had no shortage of critics. Objections to the sophists’ craft ranged from labeling them as unprincipled mercenaries who would argue any position for pay, hairsplitters who devolved into bickering over pointless nuances to score

an argumentative victory (Barilli 5), and flatterers who sought to please the crowd by deceit and trickery (*Gorgias*, 463a6-465e1). Beyond the ad hominem attacks, these criticisms proceed from a fundamental assumption about rhetoric: It only seeks to persuade (by whichever means are at its disposal), but does not pursue genuine truth.

Plato articulated this criticism in its most lasting form by casting rhetoric in an opposition with—and inferior to—philosophy. Rhetoric was “mere cookery” and “flattery” of the masses (ibid.). It could only aspire to the probable, perennially uncertain, and was thus confined to the realm of opinion (*doxa*). Only philosophy was committed to truth (*episteme*) (Barilli 6). Philosophy sought to eliminate falsehoods, perspectivism, and uncertainty by employing the method of dialectics: claims are subjected to critical questioning, and those that do not withstand it are ruled out (Toye 9-12). Rhetoric was thus a distracting overlay of stylistics and deception, an ethically and epistemologically questionable practice that had to be cleared away.

Plato’s critique of sophism rests on positing a dichotomy between truth and opinion. Truth is the domain of philosophy, which allows it to oust rhetoric and consign it to the lower realm of opinion. Philosophy’s pursuit of truth relies on a true method—dialectics—, while rhetoric strings together whichever devices it can find to “make the weaker argument appear stronger.” Rhetoric’s inability to pursue truth, in turn, reduces it to an empty formalism. Of course, Plato’s move to cast philosophy and rhetoric as a hierarchical binary is easily unmasked as a rhetorical move in its own right to serve the ends of legitimizing philosophy (Derrida 103-04). Nonetheless, Plato’s distinction between truth and opinion and rhetoric’s derivative status have persisted in various forms throughout the centuries.

3. Aristotle: Finding the Available Means of Persuasion in a Given Situation

Aristotle accepted Plato's episteme-doxa distinction, but sought to rehabilitate rhetoric by showing that it straddled the realms truth and opinion. "[R]hetoric is a combination of the science of logic and of the ethical branch of politics; and it is partly like dialectic, partly like sophistical reasoning" (*Rhetorica* I, 1359b10-12). While rhetoric cannot establish certainties, neither does it aim to persuade its audience of lies or phantasms. Instead, it argues for what is probable, usual, or factually the case (1357a14, 1357a34). The techniques of rhetoric are not quite as rigorous as scientific or dialectical demonstration, but neither are they unmoored from any standard of rationality (1355a5-10).² For example, while dialectic uses syllogisms—formal deductions from premises to conclusion—, rhetoric uses enthymemes—shortened syllogisms that leave obvious premises unstated (1357a34). Similarly, while dialectic uses induction, rhetoric reasons from a few illustrative examples. A rhetor, according to Aristotle, thus traces the outlines of logical and dialectical rigor, without following every step: a rhetor *appeals* to *logos* without fully employing it (1356a19-21). This, in turn, leaves the critical audience free to reject the rhetor's argument, since enthymeme and example need not be compelling, while the conclusion from a well-reasoned deduction or induction is inescapable.

Beyond *logos*, a rhetor attempting to persuade an audience can appeal to her own authority and credibility (*ethos*), or to the audience's beliefs and preferences (*pathos*) (1356a1-21). According to Aristotle, any of these three types of proof may persuade a listener, which occurs once the listener perceives that a proposition has been *sufficiently* demonstrated. In practice, the three types rarely appear in isolation (Campe 2008, 371-73). Given that the precise combination of logical, ethical, and pathetic appeals can vary from situation to situation,

² "Persuasion is clearly a sort of demonstration, since we are most fully persuaded when we consider a thing to have been demonstrated. The orator's demonstration is an enthymeme, and this is, in general, the most effective of the modes of persuasion. The enthymeme is a sort of syllogism, and the consideration of syllogisms of all kinds, without distinction, is the business of dialectic, either of dialectic as a whole or of one of its branches."

Aristotle defined rhetoric as “the faculty of observing in any given case the available means of persuasion” (*Rhetorica* I, 1355b26-27). The rhetorical situation also requires the rhetor to think through which arguments to choose (*inventio*), how to sequence them (*dispositio*), and which stylistic devices to choose (*elocutio*). When giving a speech, the rhetor also had to memorize it (*memoria*), and practice its delivery (*pronuntiatio*) (Barilli 13-19). These five canons have been central to classical rhetorical treatises ever since.

For our purposes, Aristotle’s conception of rhetoric thus makes four corrections to Plato: 1) Rhetoric is on a spectrum with science and philosophy, but still distinct from them; 2) rhetoric participates in the search for truth, but not with the same competency; 3) rhetoric is less formalistic as science and philosophy are; and 4) Aristotle reverses Plato’s equation of rhetoric and formalism and returns science and philosophy to the realm of the formal.

But Aristotle’s model does not merely present rhetoric as a less formalistic mode of reasoning, but also as one that constantly runs the risk of unraveling. Pathetic appeals to the audience’s emotions always risk introducing demagogic excesses into rhetoric and undo its careful balance with logical appeals (Garsten 115-19). To this, Aristotle responds that emotions should not be excluded from the deliberative process, but channeled into political institutions (Garsten 128). The task of those institutions was to tether the audience’s well-being to the outcome of the deliberation. Based on the audience’s experiences, commitments, and biases, the orator had to identify deliberative pathways that would make the audience understand the best possible outcome for them, which would be a check against unjust or misguided outcomes (*ibid.*). In the words of Bryan Garsten, rhetoric, for Aristotle required a “situated judgment” by harnessing an audience’s “deliberative partialities” (Garsten 124, 135). Formalization, for Aristotle, is thus not a mere logicism, but a process of institutionalization, of creating practices that channel, diffuse, and leave room for, contingencies.

4. Cicero: Institutionalized Adversarial Testing

While Aristotle's conception of rhetoric resists the Platonic attack that rhetoric is incapable of discovering truth and partaking in the rigors of dialectical reasoning, he nonetheless places philosophy above rhetoric (Toye 12-13). Rhetoric may lead to truth, but by truncated paths of reasoning, situationally appropriate appeals to the speaker's authority and the audience's emotions, and under institutional safeguards. Aristotle, after all, was still a philosopher.

Cicero, in turn, was a rhetor. It is thus unsurprising that he reversed the hierarchy between rhetoric and philosophy. For Cicero, the only path to philosophical truth was to test arguments in an adversarial setting. "The sole object of our discussion is by arguing on both sides to draw out and give shape to some result that may be either true or the nearest possible approximation to the truth" (quoted in Garsten 154). Ordinary opinions contained the seeds of true knowledge (Garsten 155), while rhetoric was the "art of arts" that brought this knowledge into the open (Barilli 26). The orator was thus a figure who could, if endowed with enough skill, bring rhetoric and philosophy into a harmonious union (Garsten 145). For Cicero, the attainment of truth was inextricably tied to an institutional setting in which rhetorical disputations brought rhetors and audience closer to the truth. A functioning society thus depended on political institutions that protected rhetorical discourse, widespread access to rhetorical disputations, and rhetoric as the cornerstone of a Roman citizen's education (Garsten 166-67). Rhetoric was a way of learning, living, and being for Cicero.

Unlike Aristotle, Cicero does not see rhetoric and philosophy in an uneasy alliance (Campe 2002, 138-39). Rather, rhetoric is the testing ground in which modes of formalization are proposed, tested, and adopted or discarded. Rhetoric is what peels the layers of false and imprecise opinions away from truth. Institutions are not designed to contain the excesses of

rhetoric, but to promote its virtues as an agent of philosophy. Rhetoric, guided by institutional structures, is a practice of formalization without which philosophical truths remain inaccessible.

5. Quintilian: The Science of Speaking Well

While Cicero brought rhetoric and philosophy into a symbiotic alliance, Quintilian went one step further. His twelve-volume compendium, *Institutio oratoria*, treated rhetoric as its own science—the science of speaking well (*Institutio oratoria* II.15, § 33-36)—, and gave a systematic summary and discussion of rhetorical theory up to his time. The science of speaking well applied in any realm in which the true and the false cannot be distinguished with certainty (Barilli 35-36). Drawing from numerous theories of rhetoric before him, Quintilian laid out the structural schema that a speech had to follow to be considered effective and well-done: an introduction that pithily summarizes the main facts and argument (exordium); a statement of the relevant facts (narratio); an outline of the issues and arguments the speaker wishes to make (divisio / partitio); proof of the case with facts and law (confirmatio); refutation of anticipated counterarguments (confutatio); and a conclusion summing up the case (peroratio) (Barilli 36-37). While Quintilian's work does not purport to break new ground in its understanding of rhetoric, it embodies a common approach to present the discipline: as a formalized system of categories and subdisciplines that aim at exhaustiveness. Rhetoric, in other words, is approached formalistically, ready to be repeated and entrenched.

~

This rough sketch is not intended to provide a deep historical understanding of classical theories of rhetoric, but rather to open up the problem that is the central concern for this dissertation. *Rhetoric poses a problem for formalization*. It is variously a source of deception or uncertainty; it is sometimes vilified as form without substance; it can be disruptive and

destabilizing; and on some accounts, it can be difficult to distinguish from science and philosophy. Rhetoric, in other words, places an inexorable demand on science and philosophy to recognize, contend with, or accommodate it. Framing rhetoric as an other to science and philosophy and shunting it to the margins ignores the problem, but does not render it inoperative.

This dissertation does not purport to solve the problem of formalization that rhetoric raises. Instead, it offers five case studies that do not take the rhetoric—truth dichotomy as a given, but engage the problem of formalization from different angles. The problem of formalization raises deeper questions about knowledge, ethics, and history. Before delving into the case studies, this introduction will briefly sketch how rhetoric begins to problematize those truth, morality, and history. Next, this introduction will provide a brief summary of each chapter.

III. RHETORIC AND THE PROBLEM OF FORMALIZATION: THREE DIMENSIONS

1. Rhetoric and Knowledge: Means of Figuration

Plato's attack on rhetoric as mere *doxa*, divorced from the truth of philosophy, has endured in some form or other throughout the centuries. As discussed above, this distinction has continually put rhetoric on the backfoot and generated a need to defend it. Rhetorical theory often responds that the matters with which rhetoric is commonly concerned—politics, ethics, law—are *human* affairs in which no certain knowledge is possible (Kronman 683-87). The knowledge that is best achievable in human affairs is necessarily grounded in opinion, persuasion, and agreement (Scott 1967, 9-11). Scholars have characterized this rhetorical knowledge as social knowledge (Railsback 363), cooperative deliberation (Ehninger 282-87), consensus (Farrell 1976, 1-3), practical knowledge that requires appropriate responses at the

right time (Poulakos 35-36), and a practice of finding habitation in an embodied performance (Leff 157-59), and many more.

This response may well be correct, but only by virtue of adopting the Platonic doxa-episteme dichotomy that makes rhetoric's subordinate status inevitable. However, twentieth-century theories of rhetoric have taken a more ambitious approach to staking out the domain of rhetoric. Chaim Perelman, for example, unapologetically announces that his treatise *The New Rhetoric* is intended to “constitute[] a *break with a concept of reason and reasoning due to Descartes*, which has set its mark on Western philosophy for the last three centuries” (Perelman & Olbrechts-Tyteca 1). Reason, Perelman argues, has been falsely confined to formal, deductive logic, which rests on necessary truths, is convincing to every rational being, and thus need not rely on persuasion (Perelman & Olbrechts-Tyteca 1-3). Perelman counters this epistemological paradigm by developing an elaborate theory of audience. Any proposition has to gain *adherence* by an audience (Perelman & Olbrechts-Tyteca 26-30, 45-50, 65-74). Universal scientific truths are considered convincing by the universal audience, while particular audiences can be persuaded of particular truths (Perelman & Olbrechts-Tyteca 28-31). In practice, however, the universal audience is never truly universal. Rather, any particular audience that requires no persuasion for certain propositions treats those propositions as universal truths, and the audience presumes itself to incarnate the universal audience on those occasions (Perelman & Olbrechts-Tyteca 35). The question thus becomes which argumentative techniques are available to secure an audience's adherence, to which Perelman devotes the majority of his book. For Perelman, science and deductive logic are thus a subset of argumentative techniques within the broader realm of reason—and the discipline to study all of reason's argumentative techniques is rhetoric.

Perelman's intervention to focus on argumentative techniques—*how* knowledge is arranged and established as such—recalls Aristotle's typology of logical, ethical, and pathetic appeals and the various shortcuts that rhetorical reasoning can take from the probable to persuade its audience. Indeed, Aristotle stated as much when he insisted that in "rhetoric we look upon as the power of observing the *means* of persuasion on almost any subject presented to us; that is why we say that, in its technical character, it is not concerned with any special or definite class of subjects" (*Rhetorica* I, 1355b32-36). Rhetoric is thus not to be understood as the realm of probabilistic knowledge, and thus a lesser version of science, logic, or philosophy. *Rather, rhetoric, in its most general terms, studies the figuration of knowledge:* the techniques of finding, arranging, presenting, and delivering arguments to an audience in a particular situation. Rhetoric deals with how knowledge is produced, not the product as such. Any area of knowledge (be it mathematics, economics, the law, or literature) has its own figurative principles—and thus its own rhetoric.³ To study the rhetoric of a field is thus to inquire into and confront it with the criteria of deriving, arranging, and defending its own knowledge. Studying the rhetoric of a field is to pose question (and problem) of how that field is formalized. Some fields are more receptive to such a question than others—and respond to rhetorical inquiries accordingly.

2. Rhetoric and Ethics: The Silent Call to Responsibility

In classical rhetorical theory, objections to rhetoric centered around the worry that the speaker might manipulate the audience by appealing to their emotions instead of reason (Remer 402-04). The flattering rhetor might thus deceive the audience about what is good and

³ Indeed, there are studies on all of those topics. See Reyes (mathematics), McCloskey (economics), Sarat & Kearns (law), de Man (literature and criticism).

just, which recalls Plato's critique that rhetoric does not provide access to true knowledge (Remer 405; Plato, *Gorgias* 447a-466a). Moreover, emotional appeals are ethically questionable because they risk depriving the audience of their autonomy to choose freely (even if it chooses the good and just) (Remer 406). The new rhetorics of the twentieth century have distilled these critiques into a worry about ideology: To what extent does discourse keep the citizen in a state of subjection, i.e. deception and non-autonomy? (McKerrow 92-95).

Classical defenses against the ethical criticism that rhetoric manipulates its audience come from Aristotle and Cicero. Aristotle confronts the critique by giving a cognitive account of emotions (Remer 408-10). Aristotle did not view emotions as "unthinking impulses," but as guides to reasoned judgment. In the juridical sphere, persuading the judge required to make him understand what affect the accused was in at the time of the deed, against whom the affect was directed, and for what reasons (Campe 2008, 368). The rhetor's task in court was to arouse the same affect in the judge and to lead him to reach the same judgment as the rhetor (Campe 2008, 368-69, 372-73). In the political sphere, true beliefs about our own needs and what is good for us originate from our feelings, and those feelings can be modified by rational argument (Garsten 139-41, Remer 415-16). Drawing on the participants' self interest in making deliberative judgments mitigates both the consequentialist and the autonomy critique. Feeling what one needs is to *know* what one needs, and to know so authentically. Rhetoric as affectation is thus not disconnected from rhetoric as reasoned argumentation for Aristotle.

However, as Gary Remer has pointed out, Aristotle's psychological account only partially dispenses with the worry over emotional manipulation. The audience may know what it needs, but a skillful orator who "observ[es] in any given case the available means of persuasion" (*Rhetorica* I, 1355b26-27) may trick the audience into taking up a belief that, if acted upon, does not fulfill those needs (Remer 420-21). The only check on the rhetor's

persuasive abilities are her own ethical reservations. Demagoguery is still possible. In addition, speakers who decide cling to their opinions and seek to win at all costs will never be persuaded by a well-reasoned argument (Wetters 127). As Eugene Garver put it, rhetoric for Aristotle does not provide substantive moral standards. “There is no ethics of rhetoric, there is only ethics of rhetoricians, and that is just ethics itself” (Garver 229).

Cicero’s account tried to derive ethical constraints from the art of rhetoric by presenting the audience’s emotions as a source of moral demands on the speaker. The rhetor is not free to simply influence the audience as she pleases. Instead, she has to account for the audience’s expectations, the institutionalized norms of the situation, and what is appropriate at the time, place, and context of the speech (Remer 438-41). In other words, the rhetor is bound by decorum, which, in large part, is infused with the moral values of the audience (ibid.). But while this addresses the issue of the unconstrained rhetor, it raises the problem of morally bad *audience*. Cicero is forced to shift the ethical obligation back to the rhetor who, in those situations, has to overwhelm the audience with emotional appeals to change their beliefs for the better (Remer 436). This raises the troubling question: how does one ensure that the audience is not infused with bad values—harmful ideologies—to begin with? Modern rhetorical scholarship addresses this question through its theory of rhetorical personae.

Edwin Black’s seminal article, *The Second Persona*, opens with the assertion that “[t]he moral evaluation of rhetorical discourse is a subject that receives and merits attention” (Black 109). The central theoretical challenge is to connect “features of a linguistic act” to the “characteristics of the language user” (Black 110). The theoretical construct seeking to make this connection is the second persona. The second persona is the ideal audience that the speaker has in mind when she addresses the real audience. She tries to move the real audience closer to that ideal audience by using her rhetorical skills to persuade, influence, or otherwise

affect the real audience (Black 113). When a speaker addresses an audience with the aim to persuade, move, or in any way influence it, the speaker aims to effect a change in that audience (ibid.). By looking for repeated stylistic and verbal tokens, the critic can discern the second persona from the speaker's rhetorical act, and thus subject the speaker to a moral critique (Black 111). In choosing an ideal audience and deciding to move the real audience towards it, a speaker creates character of herself—an ethos—that is open to critical assessment and moral judgment (Black 110). The character implied and created by this rhetorical act is the first persona (Black 111).

Subsequent scholars quickly built on Black by taking up was the question of who and what the first-and-second-persona model ignores, excludes, and silences. For example, a speaker can marginalize or exclude groups that should be part of the second persona. This excluded set of audiences is captured in Philip Wander's third persona. The third persona concept expands the ethical lens by bringing excluded non-subjects into view and giving a voice to their claim for moral recognition (Wander 216). Furthermore, the speaker might address an audience, but use language that can only be decoded by some. If the speaker tries to disguise himself and pass for something that he is not, some in the audience might be duped, but others might notice. That clairvoyant part of the audience is Charles Morris's fourth persona. The speaker needs the fourth persona's support and collusion in his efforts to pass (Morris 230). While he may find a safe haven of like-minded kindred spirits in them, he has to constantly be on his guard for fear of being exposed to the dupes (Morris 231). The speaker thus has to throw pink herrings—deflective rhetorical devices—into the audience, which the fourth persona has to decode (ibid.). Making explicit what is unsaid is an ethical undertaking, because it exposes to scrutiny what was previously masked.

But analyses into who is excluded from the position of audience and speaker, and what is excluded from rhetorical discourse (and for whom) raises the fundamental concern that rhetorical discourse *always* leaves *something* unsaid. Indeed, the purpose of any rhetoric—be it persuasive, epistemic, invitational, etc.—is to achieve a linguistic effect without being fully self-transparent (McKerrow 97-98). This lack of transparency is at the heart of Plato's critique that rhetoric dabbles in mere opinion, and of Aristotle's contention that, unlike dialectic, rhetoric uses shortcuts in reasoning such as enthymeme and example. Modern theorists of all stripes, too, brush up against the question of what rhetoric leaves unsaid, what is unheard, and who gets to have a say. I.A. Richards' key insight was that words and language function metaphorically, i.e. that the full meaning of what is said can only come out in the context in which words interanimate each other (Richards 28-31). J.L. Austin's central focus was on illocutionary speech act—the force with which the speaker seeks to endow an utterance in order to achieve an effect (Austin 98-107). Perelman worked the Aristotelian notion of argumentative shortcuts into a detailed account of how particular audiences find specific loci, facts and truths, values, presumptions, and special agreements persuasive in argumentation, but not others (Perelman & Olbrechts-Tyteca 65-113). And Burke derived his central concept of identification from the fact that every act of dissociation in speech contains an unsaid identification: a sameness from which the argument proceeds as a starting point to argue for distinctness (Burke 19-31).

Rhetoric's tension with ethics is thus inescapable because, unlike a mathematical or (ideal) philosophical proof, it always leaves something unsaid. If electing to say something always entails *not* saying something else, then rhetorical acts become ethical choices and acts of power, open to critique and a valuative assessment. This, of course, is *not* to say that rhetorical activity is inherently immoral. Rather, rhetoric can be used for good or for ill. The

morality of each rhetorical act is open, and the onus is on the speaker to choose well while bearing that burden. If every rhetorical act exemplifies a certain character or a set of values that the rhetor endorses, then rhetorical study not only points to moral failures in those choices, but also draws attention to the ethical burden that is placed on every speaker in every rhetorical situation.

Rhetoric thus raises a problem of formalization in the ethical realm as well. Just as rhetoric figurates knowledge without being a fixed body of knowledge, so it *disposes* us towards ethics without being a fixed set of values. A rhetorical act involves an ethical choice—whom (not) to address, what (not) to say, what position (not) to take. But rhetoric only extends a silent call to ethics without providing any guidance as to which choice to make. That burden is for the rhetor to bear.

3. Rhetoric and History: Activating the Potential of the Moment

One basic trait of rhetoric over which there is no disagreement is its situationality and practicality. This consensus runs from Aristotle's notion of rhetoric as situational judgment that identifies the means of persuasion that are appropriate in the moment, to Perelman's insistence that an audience is never universal and requires arguments that respond to its particularities, to contemporary scholarship on the rhetorical situation that emphasizes that "[r]hetorical discourse is frequently distinguished from philosophical, scientific, and artistic discourses because it is judged according to criteria of particularity, contingency, and propriety. . . . [R]hetorical discourse seeks timely and fitting action" (Porrovecchio 155). Rhetorical discourse emerges under the particularities of the moment (Farrell 1991, 189-90). Audience and rhetor generate the discourse in a performative, mutually responsive and creative practice that is never wholly predictable (Vatz 159).

Rhetoric is thus inextricable from, and cannot be understood without, its emphatically historical quality. On the one hand, it requires a deep affinity with the present moment. In the words of Michael Leff, rhetoric is the “universal activity of finding habitation only in the particular” (Leff 161). Rhetoric sits in the present by grasping the moment. But to grasp the moment means, as Bitzer put it, to grasp the “rhetorical situation as a natural context of persons, events, objects, relations, and an exigence which strongly invites utterance” (Bitzer 5). It is to understand the forces and influences of the past that have led up to this situation, and the future expectations that they have created (Vatz 159). For the rhetor, the moment acquires significance by being shot through with the past that generates expectations that call for completion and exigence that strives beyond itself (ibid.). The rhetor reads the past through the present concerns for the future (cf. Campe 2015, 46-47). To appropriately respond to the rhetorical situation is to appropriately respond to the *historical* situation.

But to grasp the moment and its historicity is not merely to be swept along with it, but to generate effects and actions that shape it. Indeed, persuasion and adherence aim at nothing other than practical effects in the world. Rhetorical practices structure and affect cognitive, affective, and behavioral patterns (Scott 1973, 94). The sophists’ recognition that “[r]hetoric is the art which seeks to capture in opportune moments that which is appropriate and attempts to suggest that which is possible” (Poulakos 36), and Aristotle’s emphasis that rhetoric is “the faculty of observing in any given case the available means of persuasion” (*Rhetorica* 1355b26-27), point to rhetoric’s capacity to open up the historical moment, to shape it without being wholly beholden to it. Rhetorical practices thus shape human experience at every moment, whether those practices are visible or not (Farrell 1991, 193).

If our experience as historical beings can be termed “rhetorical,” the problem of formalization emerges in yet another variant. Rhetorical practices constantly recognize open

possibilities and contingencies, and write and rewrite the past. Rhetoric is both in history and an agent of history, both a productive process and the product itself (Leff). The instant it recognizes and shapes a moment, it risks becoming consigned to the store house of historical stock narratives. Rhetorical practices can easily become forgotten and do their work in the background. Instead of raising the problem of how (and whether) we formalize our historical experience into coherent narratives (Farrell 1991, 208), they can devolve into rote formalisms that *efface* that very problem. The rhetoricity of human experience makes the problem of formalization inherent to its historical being.

IV. THE CHAPTERS OF THIS DISSERTATION

This dissertation offers five case studies on the problem of formalization and rhetoric's role in raising and obscuring, alleviating and exacerbating that problem. Each chapter will place different emphases on the epistemological, ethical, and historical dimensions of the problem of formalization. The first chapter covers Friedrich Nietzsche's rhetorical retort to philosophy and science. As the chapter's first part illustrates, dialectical reasoning, for Nietzsche, is not the other to rhetorics, but one of several rhetorical registers. In a performance of dialectical reasoning in the *Genealogy of Morality*, Nietzsche lays bare that the ascetic ideal that animates dialectics runs into a genuine aporia when it turns back on itself: when seeking to negate itself, it either persists because it fails to do so, or it persists in *performing* a successful self-negation. In the face of such aporias, formalized, systematic structures fail to guide.

Indeed, the chapter's second part goes on to argue that uncovering aporias in formalized systems of knowledge is not only a central impetus to Nietzsche's work, but the source of a truly ethical life. Only without the guardrails of pre-formalized imperatives can the freely philosophizing spirit devise an ethics that is truly its own and deserving of the name.

Finding an ethics in the absence of foundationalist thinking is inherently dangerous—which is why I term Nietzsche’s ethics an “ethics of danger.”

Chapter two turns to the rhetoricity of lived experience. Edmund Husserl, in his late work *The Crisis of the European Sciences* develops the lifeworld as a theoretical construct to elucidate how human experience became divorced from scientific thought. Husserl argues that, since Galileo and Descartes, science has taken an attitude towards the world that is deductive. Nature is treated as a gigantic structure that can be measured, understood, and formalized in a calculus that proceeds from general principles to particular theorems. Human experience, however, is inherently inductive, proceeding from the particular to the general. The operativity of the lifeworld is thus thoroughly inductive.

In laying out this theory, however, Husserl refines and complicates induction as an inherently intentional, historical practice. Always anticipating that the incomplete experience of the moment will be expanded, completed, or corrected by future experience, human subjectivity is both inherently historical and scientific. In charting its path, the human subject is not guided by a steady Platonic idea. Instead, the inductive operativity of human experience relies on its capacity for Aristotelian *epagoge*, i.e. its impulse to take a particular as exemplary for some universal that lies beyond itself—without knowing what universal may be, or whether it truly exists at all. So long as human experience proceeds in this inductive manner, science remains connected to life. But once the accumulation of experiences and data deceives the subject to change its attitude and believe that knowledge can be deduced from a calculus, it is no longer open to experiential input. In that moment, lively formalization degenerates into lifeless formalism, and science divorces itself from history and experience. Husserl’s theory of the lifeworld thus proceeds from a very similar concern as Nietzsche—the false allure of

scientism. By the same token, Husserl's philosophy can be viewed as a response to the dangerous life that Nietzsche's philosophy proposes.

The third chapter turns to Kafka's writings and the problem of formalization as it appears in intersubjective relations. In his *Letter to the Father*, Kafka does not do what he at first purports. Instead of justifying himself, he attacks. Instead of proving his innocence, he proves his father's guilt. Instead of freeing himself from his father's grip, he draws the two inextricably close. Kafka pursues all of these goals in a remarkably formalistic manner. Indeed, this chapter uses the logical method of gödel numbering to trace how Kafka sets up a logical proof that is designed to defeat itself and invite a response. In purporting to justify himself, Kafka weaves reproaches (which resemble those that his father used against him) into the letter to coax his father into justifying *himself*. In tempting the father to adopt the rhetorical register of guilt, Kafka seeks to level the playing field between the two: In arguing over who is at fault, the outcome of the debate is secondary to the fact that both have to speak the same language. The letter is thus an attempt by Kafka to re-formalize their relationship—and to make that relationship an engine for further writing.

The fourth chapter takes a broader view on the paradoxes of intersubjectivity. It dives into Jacques Lacan's early and middle writings in which he wrestled with, and ultimately abandoned, his attempts to formalize psychoanalysis into a science. The chapter first discusses how, in his early seminars, Lacan lays out how theories of intersubjectivity run into an infinite regress problem if they only conceptualize small "others" who attempt to divine what each of them thinks that the other thinks what they think . . . , and so forth. To break this infinite regress, Lacan accesses and theorizes a symbolic order of play that always already structures intersubjective interactions. Formalization under the symbol of the Other is always already underway. This formalization, however, is by no means a rote process. The second part of the

chapter discusses how empty signifiers both disrupt the symbolic chain and impel it. Using the example of language functions that appear like mathematical algorithms (but are no such thing) that demand to be taken literally, Lacan's writings demonstrate and perform the formalization process in which we constantly engage: to obscure the lack of a signified to a signifier, we generate more and more signifiers that we demand be taken literally, i.e. as the signified. As this never materializes, we continue to produce signifiers.

Chapter five takes a step beyond European writers of the early twentieth century and looks at the problem of formalization in American law. Taking qualified immunity⁴ as a case study, this chapter looks at how federal courts allocate power between the legislative, executive, and judicial branches, as well as between governmental institutions and the people. The chapter develops the figure of the "Marburian move," by which courts avoid taking a side in the power politics of private citizen invoking the law to hold a government official accountable for violating the law. In a move that dates back to the iconic case *Marbury v. Madison*, a court declines the legislature's grant of power to adjudicate a class of cases against government officials. In doing so, the court hands a small victory to the government, denies legal relief to the private citizen, while insisting that the court could hold the government official accountable if it had the jurisdiction to do so.

This move takes a middle position between two extremes that gives courts room to maneuver in individual cases and develop a more intricate judicial doctrine. The judiciary thereby expands its own domain of power in the American constitutional structure by formalizing doctrine that maintains flexibility. Yet, over time, qualified immunity doctrine

⁴ Qualified immunity can be invoked by government officials who are sued for violations of federal rights. In essence, qualified immunity creates extra room for error for those officials, compared to private citizens. Only if the right in question is clearly established can the lawsuit go forward.

turned from a source of flexibility into a rote formalism that, performing the same legal analysis again and again, consistently favored government officials. In abandoning the flexibility that the Marburian move once afforded it, the Supreme Court has come under political pressure for eviscerating individual rights and giving a *carte blanche* to government officials abusing their power.

Lastly, it should be noted that chapter five deliberately uses the Bluebook citation style, while the first four use the MLA style. This is because much the source material on which the fifth chapter relies—statutes, cases, appellate briefs, memoranda—have a straightforward way to be cited in the Bluebook style, but not in others. At the same time, I wanted to use a citation style that is common to graduate departments wherever possible. Since the citation style is not uniform across all chapters, I opted against a bibliography for the dissertation, and for chapter-specific “Works Cited” sections.

WORKS CITED

- Aristotle. *Rhetorica*. In: Richard McKeon (ed.) *The Basic Works of Aristotle*. New York: The Modern Library, 2001. 1325-1451. Print.
- Austin, J. L. *How to Do Things With Words*. ed. J. O. Urmson. Cambridge, MA: Harvard University Press, 1962. Print.
- Barilli, Renato. *Rhetoric*. trans. Giuliana Menozzi. Minneapolis: University of Minnesota Press, 1989. Print.
- Bitzer, Lloyd F. "The Rhetorical Situation." *Philosophy and Rhetoric* 1.1 (1968). 1-14. Print.
- Black, Edwin. "The Second Persona." *Quarterly Journal of Speech* 56.2 (1970). 109-119. Print.
- Burke, Kenneth. *A Rhetoric of Motives*. Berkeley: University of California Press, 1969. Print.
- Campe, Rüdiger. *Spiel der Wahrscheinlichkeit: Literatur und Berechnung zwischen Pascal und Kleist*. Göttingen: Wallstein, 2002. Print.
- — —. "An Outline for a Critical History of *Fürsprache*: *Synegoria* and Advocacy." *Deutsche Vierteljahresschrift für Literaturwissenschaft und Geistesgeschichte* 82 (2008). 355-381. Print.
- — —. "Rhetoric's Flirtation with Literature, from Gorgias to Aristotle: The Epideictic Genre." In: Daniel Hoffman-Schwartz, Barbara Natalie Nagel, Lauren Shizuko Stone, eds., *Flirtations: Rhetoric and Aesthetics This Side of Seduction*. New York: Fordham University Press, 2015. 37-50. Print.
- Conley, Thomas M. *Rhetoric in the European Tradition*. New York: Longman, 1990. Print.
- de Man, Paul. *Blindness and Insight: Essays in the Rhetoric of Contemporary Criticism*. 2nd rev. ed., intro. Wlad Godzich. Minneapolis: University of Minnesota Press, 1983. Print.
- Derrida, Jacques. "The Pharmakon." In: Jacques Derrida. *Dissemination*, trans., ed. Barbara Johnson. Chicago: University of Chicago Press, 1981. 95-117. Print.
- Ehninger, Douglas. "Decision by Debate: A Re-Examination." *Quarterly Journal of Speech* 45 (1959). 282-287. Print.
- Farrell, Thomas B. "Knowledge, Consensus, and Rhetorical Theory." *Quarterly Journal of Speech* 62.1 (1976). 1-14. Print.
- — —. "Practicing the Arts of Rhetoric: Tradition and Invention." *Philosophy and Rhetoric* 24.3 (1991). 183-212. Print.
- Gagarin, Michael. *Antiphanes the Athenian: Law and Oratory in the Age of the Sophists*. Austin: University of Texas Press, 2002. Print.

- — —. “Rhetoric and Law.” In: Michael J. MacDonald, ed. *The Oxford Handbook of Rhetorical Studies*. Oxford: Oxford University Press (2017). 43-52. Print.
- Garsten, Bryan. *Saving Persuasion: A Defense of Rhetoric and Judgment*. Cambridge, MA: Harvard University Press, 2009. Print.
- Garver, Eugene. *Aristotle’s Rhetoric: An Art of Character*. Chicago: University of Chicago Press, 1995. Print.
- Goodrich, Peter. “Rhetoric and Law.” In: *The Oxford Handbook of Rhetorical Studies*. 613-624. Print.
- Hariman, Robert. “Status, Marginality, and Rhetorical Theory.” *Quarterly Journal of Speech* 72.1 (1986). 38-54. Print.
- Hollander, John. “Legal Rhetoric.” In: Peter Brooks & Paul Gewirtz, eds. *Law’s Stories: Narrative and Rhetoric in the Law*. New Haven: Yale University Press, 1996. 176-186. Print.
- Kennedy, George. “A Hoot in the Dark: The Evolution of General Rhetoric.” *Philosophy and Rhetoric* 25.1 (1992). 1-21. Print.
- Kronman, Anthony T. “Rhetoric.” *University of Cincinnati Law Review* 67 (1999). 677-709. Print.
- Leff, Michael C. “The Habitation of Rhetoric.” In: Antonio de Velasco, John Angus Campbell, David Henry (eds.) *Rethinking Rhetorical Theory, Criticism, and Pedagogy: The Living Art of Michael C. Leff*. Ann Arbor: Michigan State University Press, 2016. 143-162. Print.
- McCloskey, Deirdre N. *The Rhetoric of Economics*. 2nd ed. Madison: University of Wisconsin Press, 1998. Print.
- McKerrow, Raymie E. “Critical Rhetoric: Theory and Praxis.” *Communication Monographs* 56.2 (1989). 91-111. Print.
- Morris III, Charles E. “Pink Herring and the Fourth Persona: J. Edgar Hoover’s Sex Crime Panic.” *Quarterly Journal of Speech* 88 (2002) 228-244. Print.
- Quintilian. *Institutio oratoria*. trans. Harold Edgeworth Butler.
<https://www.perseus.tufts.edu/hopper/text?doc=Perseus%3Atext%3A2007.01.0060%3Abook%3D2>. Web.
- Reyes, G. Mitchell. “The Rhetoric in Mathematics: Newton, Leibniz, the Calculus, and the Rhetorical Force of the Infinitesimal.” *Quarterly Journal of Speech* 90.2 (2004). 163-188. Print.
- Richards, I.A. *The Philosophy of Rhetoric*. New York: Oxford University Press, 1965 [1937]. Print.

- Sarat, Austin & Kearns, Thomas R., eds. *The Rhetoric of Law*. Ann Arbor: University of Michigan Press, 1996. Print.
- Scott, Robert L. "On Viewing Rhetoric as Epistemic." *Communication Studies* 18.1 (1967). 9-17. Print.
- — —. "On *Not* Defining Rhetoric." *Philosophy and Rhetoric* 6.2 (1973). 81-96. Print.
- Parker, Douglas H. "Rhetoric, Ethics, and Manipulation." *Philosophy and Rhetoric* 5.2 (1972). 69-87. Print.
- Perelman, Chaim & Olbrechts-Tyteca, C. *The New Rhetoric: A Treatise on Argumentation*. Notre Dame: Notre Dame University Press, 1971. Print.
- Plato, *Gorgias*. In: John M. Cooper & D. S. Hutchinson, eds. *Plato. Complete Works*. Cambridge, UK: Hackett, 1997. 791-869. Print.
- Porrovecchio, Mark J., "The Evolution of the Rhetorical Situation." In: Mark J. Porrovecchio & Celeste Michelle Condit, eds. *Contemporary Rhetorical Theory: A Reader* (second ed.). New York: Guilford Press, 2016. 155-159. Print.
- Porrovecchio, Mark J. & Condit, Celeste Michelle. "What Can a 'Rhetoric' Be?" In: *Contemporary Rhetorical Theory: A Reader*. 15-20. Print.
- Poulakos, John. "Toward a Sophistic Definition of Rhetoric." *Philosophy and Rhetoric* 16.1 (1983). 35-48. Print.
- Railsback, Celeste Condit. "Beyond Rhetorical Relativism: A Structural-Material Model of Truth and Objective Reality." *Quarterly Journal of Speech* 69.4 (1983). 351-363. Print.
- Remer, Gary. "Rhetoric, Emotional Manipulation, and Political Morality: The Modern Relevance of Cicero vis-à-vis Aristotle." *Rhetorica* 31.4 (2013). 402-443. Print.
- Toye, Richard. *Rhetoric: A Very Short Introduction*. Oxford: Oxford University Press, 2013. Print.
- Vatz, Richard E. "The Myth of the Rhetorical Situation." *Philosophy and Rhetoric* 6.2 (1973). 154-161. Print.
- Vickers, Brian. *In Defence of Rhetoric*. Oxford: Oxford University Press, 1988. Print.
- Wander, Philip. "The Third Persona: An Ideological Turn in Rhetorical Theory." *Communication Studies* 35.4 (1984) 197-216. Print.
- Wetters, Kirk. *The Opinion System: Impasses of the Public Sphere from Hobbes to Habermas*. New York: Fordham University Press, 2008. Print.

CHAPTER 1: NIETZSCHE: RHETORIC AGAINST A FORMALIST WAY OF LIFE

Nietzsche distrusted logics—philosophical or scientific—that laid a totalizing claim to deriving true knowledge or universally binding ethical precepts. Instead, this chapter argues, Nietzsche viewed those logics as operating in a larger rhetorical matrix. Persuasive rhetorics take on the veneer of inexorable logic once they are able to efface their situatedness and present themselves as the only plausible mode of reasoning. The first part of this chapter illustrates this thesis in a case study of Nietzsche's view on Hegelian dialectics. It compares Nietzsche's *On the Genealogy of Morality* to the paradigmatically Hegelian *A Contribution to the Critique of Hegel's Philosophy of Right* by Karl Marx. Marx sees dialectics as the inexorable logic of history. Yet, he has to rely on the rhetorical persuasiveness of the chiasmus to make his claim plausible. Nietzsche, on the other hand, conceives of two incompatible logics: the nobles' positive affirmation (non-dialectical) and the priests' negative oppositionality (enabling dialectics). Instead of arguing for one logic over another, Nietzsche foregrounds their rhetoricity by performing the historically contingent invigoration and desiccation of each, leaving it to the reader to assimilate whichever mode of philosophizing they find most plausible. At the end of the *Genealogy*, Nietzsche confronts the reader with a paradox: does the ascetic spirit's attempt to negate itself result in its self-overcoming, or does it survive in performing the gesture of self-negation? Nietzsche says nothing on how to resolve the paradox. Instead, he falls silent.

The second part of this chapter studies the self-negating, paradox-seeking impulse across Nietzsche's oeuvre and studies its ethical implications. Danger, this chapter argues, is a motif that provides a fulcrum between Nietzsche's paradoxical, self-undermining reasoning and its unconventional ethics. Danger is the threat of complacent thinking and entrenched belief systems disintegrating under their hidden self-contradictions. It emerges as a key

concern in Nietzsche's early works and brings with it a persistent challenge to our thinking and our ethics; refusing to be deceived, we must live courageously by identifying the paradoxes that beset our knowledge and moral beliefs. In his middle works, Nietzsche takes on the persona of the intellectually honest free spirit that attacks any systematic metaphysics, and thus offers a dangerous proposition: An ethics is only possible (yet may still fall short) in moments when metaphysics collapses and paradox rears its head.

PART I: THE RHETORICAL PERFORMANCE OF PHILOSOPHY⁵

1. The Offense of Hegelianism

In his stock-taking work *Ecce Homo* of 1888, Nietzsche identifies a major grievance with his youthful debut publication, *The Birth of Tragedy*: “it smells offensively Hegelian [...] One ‘idea’ – the opposition between Dionysian and Apollonian – translated into metaphysics; history itself as the development of that ‘idea’; the opposition sublated into a unity in tragedy” (*EH*: 108). Indeed, Hegelianism was a *bête noire* of which Nietzsche strove to steer clear for the rest of his intellectual career. In the second *Untimely Meditation*, he asserts: “I believe that there has been no dangerous vacillation or crisis of German culture this century that has not been rendered more dangerous by the enormous and still continuing influence of this philosophy, the Hegelian” (*HL*: §8, 104). Nietzsche particularly rejected the Hegelian philosophy of history, according to which the inexorable forces of opposition compel an objective spirit to progressively reveal itself (*D: P* §3, 4; *BGE*: §244) and sublate those oppositions into a higher, harmonious, utopian, state in which history comes to an end (*HL*:

⁵ This part was originally published as Kuchne, Tobias. “Nietzsche and the Rhetoric of Dialectics.” *Journal of European Studies* 48.2 (2018). 115-132. Print. Author retained the right to reprint the article in any book written by the author.

§8, 104).⁶ Particularity and difference will not, according to Nietzsche, eventually be subsumed into an overarching, abstract universalism (*CW*: §10, 252). Nietzsche recoils at such a Hegelian philosophy of history because it would enshrine the present as an inevitable outcome of the past, making us mere epigones without historical agency (*HL*: §8, 104). In other words, Nietzsche rejects Hegelianism for its teleological (and theological) ethos of thinking history from its end, instead of giving the present moment access to an open future.

From this perspective, it is not surprising that Nietzsche never directly engages with authoritative representatives of Hegelianism such as Marx, or Hegel himself. Nietzsche was not interested in continuing their legacy through a recuperative or corrective critique, but was content with polemically attacking and dismissing lesser representatives of “the most infamous of all corrupters of German, the Hegelians and their deformed offspring” (*DS*: §12, 54) in order to develop his own philosophical style. As such, Nietzsche makes examples of David Strauss, whose inchoate thinking is a result of the disease of “Hegelism and Schleiermachism” that he contracted in his youth (*DS*: §6, 27), Eduard von Hartmann, who derived a laughable philosophy of history from a shallow Hegelianism (*HL*: §9, 104 and 108), Eugen Dühring, who put a coarse set of oppositions in the service of a petty anti-Semitism (*GM*: III.14, 91), and Richard Wagner, who, having been led astray by Hegel in his youth (*GS*: §99, 153), later committed the worst possible offense against art: “he invented a style that ‘meant the infinite’, – he became *Hegel’s heir* ... Music as ‘Idea’ –” (*CW*: §10, 252). Nietzsche used the label “Hegelianism” as an insult. For him, it was equivalent to thinking in facile oppositions that could naively be resolved into a higher unity: “Hegelianism” was a synonym for sloppy,

⁶ While there is still disagreement over the exact reconstruction of Hegel’s dialectical method, it is clear that Nietzsche takes issue with its most fundamental assumptions, as laid out by Forster describes them in the *Cambridge Companion to Hegel*: that thought (in whichever manifestation) can recognize its finitude and contradictions, sublate itself and transition into their opposites, and ultimately resolve and overcome them in a higher unity (Forster 1993, 132).

unsophisticated philosophizing *as such*. Instead of a recuperative or corrective critique of Hegelian thought, Nietzsche was interested in an altogether different way of thinking that would allow for an escape from the sickening nihilism of his time.

In spite of Nietzsche's fierce disdain for Hegelianism, a puzzling trend persists in Nietzsche scholarship. Authoritative readers of Nietzsche have persistently sought to align him with Hegel. Jaspers insists that, in order to understand Nietzsche's method of philosophizing, we have "to bring his thoughts into juxtaposition and enter into the *real dialectic* through which alone his intention becomes clear" (Jaspers 11). In his seminal *Nietzsche: Philosopher, Psychologist, Antichrist*, Walter Kaufmann identifies Nietzsche as a "dialectical monist" (Kaufman 235-46) whose notion of will to power functions in a similar way to the Hegelian spirit's *Aufhebung*: "Nietzsche's sublimation actually involves, no less than does Hegel's *aufheben*, a simultaneous preserving, canceling, and lifting up" (236, see also 132-33, 408). Zimmermann continued Kaufmann's line of thought by arguing that Nietzsche "has an Hegelian, which is to say, dialectical notion of negation" (Zimmermann 277) and that "Nietzsche is an Hegelian" who sees history progress to the advent of a "fully realised and developed, self-made man, the *Übermensch*" (280).

Breazeale and Houlgate cautioned against the urge to read a dialectical method into Nietzsche's philosophy (Breazeale 149; Houlgate 13-15). Houlgate, Dudley, and Williams continued a more modest line of research by placing Hegel and Nietzsche in a shared project of a post-Kantian critique of metaphysics (Houlgate), freedom (Dudley), and morality (Williams). Most recently, however, scholarship has thrown this caution to the winds again. In her reading of *Zarathustra*, Cauchi asserts that "Nietzsche engages in a discipline of thought that proceeds on the dialectical principle of affirmative negation, and in doing so recuperates Hegel's ... 'labour of the negative'" (Cauchi 2). She thus seeks to revive Zimmermann's

ambitious argument that the *Übermensch* stands at the end of a progressive history of self-overcoming that follows the path of dialectics. And yet, she is forced to conclude her article with the admission that the Hegelian dialectic is rehearsed but never achieved:

Whereas in the Hegelian dialectic, ‘the positive elements in which [the] negative element disappears as something subordinate and vanquished’ (1953: 18) prevail as signifiers of a predetermined plan, in the Zarathustran dialectic, the negated elements reassert their dominance in a vicious cycle of eternal return. [...] [T]he existential reality is a relentless labour of the negative without the redeeming positive of progression. (Cauchi 14)

Despite Nietzsche’s open aversion to Hegelianism, scholarship has repeatedly felt the temptation to find affinities between them.

The disagreement among scholars who pursue this line of inquiry is complicated by a wider panoply of voices that insist on a more ambivalent, non-Hegelian logic,⁷ and those who propose to find access to Nietzsche’s philosophy from an altogether different angle: his rhetoric and style.⁸ This vast divergence of approaches is symptomatic of the fundamental question posed by Breazeale in his stock-taking article, entitled “The Hegel-Nietzsche Problem,” but has so far been sidestepped: “To this day, the most important unresolved issue [revolves around] the problem confronted by Hegel and Nietzsche: what is the relationship between affirmation and negation?” (1975: 162). This part seeks to tackle this more underlying question, not only to clarify whether Nietzsche can be meaningfully called a dialectician, but

⁷ Deleuze’s *Nietzsche and Philosophy* is commonly cited as representing the view of Nietzsche as an anti-dialectician, although it is now discredited for polemically misrepresenting Hegel’s dialectics and thus Nietzsche’s divergence from it (see Cauchi 2, Houlgate 7-8, Breazeale 161). In a similarly non-Hegelian vein, Müller-Lauter’s influential account of Nietzsche’s “philosophy of contradictions” traces how Nietzsche, despite his best efforts, repeatedly fails to reconcile the oppositions and contradictions in his thinking (Müller-Lauter 5). Other interpreters who have insisted on a non-Hegelian logic in Nietzsche are Zitko (1992), Zittel (1995), and de Miranda (2006).

⁸ In the French tradition, Derrida (1979 [1978]) and de Man (1979a and 1979b) have most famously espoused this view. In the English-speaking world, Nehamas pursues a similar thought by suggesting that the tensions in Nietzsche’s are resolved once they writings are viewed as literature instead of philosophy (Nehamas 3). More focused inquiries into the import of rhetoric and style in Nietzsche were made in the German-speaking *Nietzsche-Studien*. A few examples are Robling (1996), Renzi (1997), Behler (1998), and Simonis (2002).

also what it means to philosophize in a Nietzschean sense: for Nietzsche, negation and affirmation are not self-evident logical operations, but are activated in rhetorical contexts that cannot be ignored. Nietzsche is neither a rigid systematizer, nor a poet-turned-philosopher. Rather, he alerts us to, and takes seriously, the fact that any kind of philosophizing (including the dialectical), has to be interrogated within its rhetorical matrix. This is where Nietzsche's philosophy exceeds the Hegelian.

2. Marx: The Chiasmic Structure of Dialectics

To shed light on how affirmation and negation function in Nietzsche's thought, and whether it can be meaningfully described as dialectical, it is helpful to start from an indisputable example of dialectical thinking, assess how affirmation and negation intertwine in it, and see how Nietzsche treats them by comparison. To this end, this chapter will examine two exemplary and mutually illuminating texts: *On the Genealogy of Morality* and Marx's introduction to the *Contribution to the Critique of Hegel's Philosophy of Right*.

At first glance, this choice may be surprising, given that Nietzsche steered clear of Marx even more than he did of Hegel, and that it has been unclear until Brobjer's recent article (2002) whether Nietzsche was even familiar with Marx's thought and writings. However, the juxtaposition of the two texts is warranted for a number of reasons. To begin with, both authors react to the macro-historical trend of Christianity and bourgeois society's decline. Löwith's claim in *From Hegel to Nietzsche*: "Nietzsche is the only man, after Marx and Kierkegaard, who made the decline of the bourgeois-Christian world the theme of such a fundamental analysis" (Löwith 176) holds especially true for the *Genealogy*.

Moreover, both thinkers confront this problem during a time in which Hegelianism was ubiquitous and impossible to ignore, but had degenerated into intellectually barren

epigonism. In a foreword to *Capital I*, Marx indicates that, in his *Critique of Hegel's Philosophy of Right*, he had performed a corrective critique of Hegel that salvaged Hegel's philosophy of history at a time when it had become an abstract mental acrobatics without practical efficacy (Marx 1906 [1872], 25). Indeed, the introduction to the *Critique* features the core themes that Nietzsche saw in Hegelian thought: oppositions driving history forward (Marx 1975 [1843], 246-7), a self-clarifying and self-revealing historical reason (Marx 1975 [1843], 252-53), and a sublation of contradictions in a universally realized utopian state (Marx 1975 [1843]: 254-47).⁹ In 1843, Marx thus takes *A Contribution to the Critique of Hegel's Philosophy of Right* as an occasion to revitalize Hegelian thinking, make it a motor for political action, and propel history forward in an age of disillusioned nihilism.

In 1887, Nietzsche reacts to a similar historical situation of nihilism and Hegelian epigonism, albeit in a very different way and from a different perspective. While Marx advocates for a political revolution that would later occur but fail, Nietzsche's *Genealogy* tells a story from the viewpoint of a *moral* revolution that has already succeeded: "*the slaves' revolt in morality*" (GM: I.7, 18).¹⁰ If it is truly possible to impute anything like a dialectic to Nietzsche's genealogical account of history, it is crucial to parse how affirmation and negation relate to each other in an indisputably dialectical account such as Marx's (see Holz 340). The apparent parallelisms and more deep-seated divergences between Nietzsche and Marx will provide a new angle that shows how Nietzsche could *appear* to be a dialectician, but ultimately espouses a far more expansive understanding of what it means to philosophize: affirmation and negation

⁹ Hegel and Marx's differences on politics, theology, and metaphysics are less relevant because they concern substantive disagreements and matters of "application" of the dialectic. In order to assess whether Nietzsche's thinking follows dialectical principles, the roles of negation and of history in his philosophy are of prime importance. Cf. Allen W. Wood's "Hegel and Marxism" in the *Cambridge Companion to Hegel* (Wood 416, 428).

¹⁰ Throughout the chapter, emphases are in the texts quoted, unless otherwise noted.

only reveal their full complexity and subtlety if they are considered in the context of their rhetorical deployment.

Three clear-cut oppositions structure Marx's account: religion deceives and oppresses the human subject (Marx 1975 [1843], 243); abstract philosophical thinking thwarts political practice (Marx 1975 [1843], 249); and the bourgeoisie exploits the proletariat: "If one class [*Stand*] is to be the class of liberation *par excellence*, then another class must be the class of overt oppression" (Marx 1975 [1843], 254, German in the quoted translation). All three oppositions align neatly with Marx's principal dichotomy between ideological superstructure and material base. Religion and abstract philosophy have, according to Marx, served as mere ideologies so far, which are critical for the bourgeoisie to survive (Marx 1975 [1843], 243-44, 250). On the other side of the principal divide, the proletariat, political practice, and the oppressed human subject are all firmly rooted in the reality of material conditions. Sorting all oppositions along this central division is the precondition that allows for their dialectical sublation: when consciousness apprehends the glaring contradiction between fictitious ideology and material reality, the revolution begins and abolishes religion, abstract philosophizing, and bourgeoisie in one stroke (Marx 1975 [1843], 256-57). The proletariat (and humanity as such) assumes its proper position once it grasps the role it plays in the dialectic of world history.

The dialectical sublation of historical contradictions seems to follow a rigid logical pattern for Marx: there exists a hierarchical, exploitative, and hence contradictory relation between an oppressor and an oppressed. The former reaps the benefits while the latter is exploited in procuring those benefits. Once this contradiction is recognized, it is overcome by means of a twofold negation. The terms of the contradictory relation are reversed (religion no longer determines the human, but vice versa; philosophical thought no longer obviates practice but serves its material needs; the proletariat seizes the means of production from the

bourgeoisie), and, through that reversal, the relation is overcome: the new top term of the hierarchy universalizes itself as the bottom term disappears altogether (through the revolution, the proletariat becomes the universal class). In Marx's historico-dialectical model, negation and affirmation interrelate by means of negating a negation (reversing a contradiction), which effects a positive, affirmative result (overcoming the contradiction in a stable, universal state).

However, this analysis exposes an ambiguous and opaque moment in Marx's dialectic. First, a hierarchy is negated by manipulating the terms in the relation but keeping them, before the hierarchical relation *as such* is negated by discarding one of the terms, while the other one is wholly transformed. Marx's model merges a moment that is logically plausible with one that is logically suspect: why is a reversal of terms tacitly assumed to lead, without further elaboration, to the universalization of one of them? Yet, it is the latter moment that is crucial in bringing about Marx's utopian, post-Capitalist society. The logical structure of this simultaneous reversal and overcoming is far from plausible, but the fact that it is taken to be such reveals that Marx's text has to rely on an altogether different technique to drive home its point: rhetorical persuasion.

Indeed, the rhetorical dimension of Marx's text becomes clear once we recognize that, far beyond making a merely descriptive point in the register of a logical, and hence inevitable course of history, *A Contribution to the Critique of Hegel's Philosophy of Right* seeks to intervene and become operative within its own historical context. While the Young Hegelians had been quibbling over Hegel's failure to reconcile thought and reality in religion (Marx 1975 [1843], 243-44), Marx moves beyond that purely academic debate by exposing how, more fundamentally, Hegel had failed to recognize the concrete material reality from which thinking has to proceed. Marx corrects Hegel in the pithy chiasmus: "*Man makes religion*, religion does not make man" (Marx 1975 [1843], 244). Philosophy dialectically overcomes itself and turns

into practice when it exposes the contradiction between thought and reality, grasping in that very moment that only radical, revolutionary action can sublimate that contradiction (Marx 1975 [1843], 255-56). The agent of this practice is the proletariat, which, in that moment of realization, assumes its role as the universal class. Marx concludes: “Philosophy cannot realize itself without the transcendence [*Aufhebung*] of the proletariat, and the proletariat cannot transcend itself without the realization [*Verwirklichung*] of philosophy” (Marx 1975 [1843], 257, German in the quoted translation). Indeed, Marx historically positions his text at the brink of a Hegelian moment of sublation: the promise of his teleological philosophy of history is about to be fulfilled as the text urges its readers toward a shift from thought into revolutionary action.

The chiasmus is Marx’s indispensable rhetorical vehicle in this political undertaking. Marx constantly relies on chiasmi to sum up his points, arrest the reader’s attention, and create his most convincing moments: “It is not enough that thought should strive to realize itself; reality must itself strive towards thought” (Marx 1975 [1843], 252); “In France partial emancipation is the basis of universal emancipation. In Germany universal emancipation is the *conditio sine qua non* of any partial emancipation” (Marx 1975 [1843], 255). With the chiasmic figure of “not $A \rightarrow B$, but $B \rightarrow A$,” Marx uncovers (or postures to uncover) the old false opposition “ $A \rightarrow B$ ” by reversing it into the correct relation “ $B \rightarrow A$ ”. The moment of revealing an old contradiction coincides with its correction and overcoming. Marx achieves this double effect by rhetorically setting up an opposition of oppositions (“A versus B” versus “B versus A”) in which the two appear to be on an unequal footing: Religion does not make man — man makes religion! The new opposition seeks to overcome the old one by virtue of its intuitive and immediate plausibility. Marx harnesses the architecture of the chiasmus of opposing two oppositions to make one of them seem absurd, the other self-evident. Rhetoric

postures as indisputable logic: in contrasting two oppositions, it becomes “obvious” that the new opposition is correct, the old one obsolete.

Once the new opposition has replaced the old, it is poised to vanish altogether: if humanity makes religion (and not the other way round), there is no more strife between the two; if the proletariat overturns its oppression at the hands of the bourgeoisie, it assumes its role as the universal class and overcomes class struggle altogether. The chiasmus not only shows a dialectics *in* action, but also seeks to put it *into* action by posturing as a logical structure and rhetorically infusing it with persuasive efficacy. Drawing *on* the chiasmus, Marx’s text aims to function *as* a chiasmus to rehearse, anticipate, and inaugurate a political revolution—the *turning around* of how things currently are. The chiasmus is the rhetorical motor of Marx’s dialectics, propelling us from thought to action, from a contradictory past into a utopian future.

If even a paradigmatically dialectical thinker such as Marx relies on rhetorical support for his logic to function – if the transition from negation to affirmation requires a chiasmus’ persuasive push for him –, it is clear that we need to approach the question of Nietzsche’s dialectics on an even more fundamental level than Breazeale did. Instead of asking what the relationship between affirmation and negation is, we need to question their straightforward appearance as logical operators and ask: what are the rhetorical means by which affirmation and negation become activated and persuasive? What does it mean to affirm and negate, i.e. what are the rhetorical modes in which negation and affirmation can be plausibly (or implausibly) performed and transformed into each other? To what extent can philosophizing be separated from rhetorical persuasion?

3. Nietzsche and the Historicity of Dialectics

Nietzsche's *Genealogy of Morality* revolves around a very different kind of revolution: the slaves' revolt in morality. Their chiasmic battle cry is this: "It is false that the nobles are good and, hence, that we, the priests and slaves, are bad. Rather, the nobles are evil and, hence, we are good!" Far from trying to put this assertion into practice and making it gain currency, however, Nietzsche is interested in its emergence and aftermath. He is convinced that this revolution has already taken place and was resoundingly (and lamentably) successful. The *Genealogy* seeks to understand how it was possible for this chiasmus to be conceived, instituted, and become a universal illness of thinking. Unlike Marx's text, the *Genealogy* does not attempt to utilize the chiasmus as a strategic weapon, but to unmask how one party— the slaves and priests — deployed it against another — the nobles — in a historical struggle.

This struggle is indeed one between two fundamentally different logics, i.e. two diverging conceptions of how affirmation and negation interrelate. The nobles give primacy to positive affirmation: in spontaneous effusions of strength, energy, and joyful disdain, they posit (indeed celebrate) their actions and character traits as good (*GM*: I.5 and I.10) and thereby dominate the slaves and priests, who are simply too weak to engage in the same positive activity. They receive the label "bad," not because they are morally inferior, but because of their unfortunate lack of good traits. The weak ones hardly cross the nobles' minds: "When the noble method of valuation makes a mistake and sins against reality, this happens in relation to the sphere with which it is *not* sufficiently familiar" (*GM*: I.10, 20). As such, the nobles do not possess a clear sense for distinctions, oppositions, or of negation.

A logic of negative oppositionality only appears with the priestly dichotomy of evil versus good. The priests turn the nobles' values into their opposite by reasoning: "the nobles' domineering behavior against us is evil. *Therefore*, we, the priests and slaves, are good." Priestly thinking's modus operandi is *ressentiment*: it does not posit, but negates and reverses what has

been posited by someone else (*GM*: I.10, 20). It is *eo ipso* reactive to an opponent and defines itself through what it is not. While negation is a mere afterthought in the affirmative logic of the nobles, the priests and slaves place a premium on negation: the nobles simply happen to negate because they affirm themselves, while the priests and slaves must affirm themselves through a negation of the other. The dominated priests and slaves have to tip the scales in their favor by relying on the rhetorical power of simultaneous reversal and overcoming. In declaring the nobles evil, the priests and slaves lay claim to a universal, supposedly intuitive truth: if the nobles are morally evil, then we (the priests) must be morally good! The priests' path to power is the performance and institution of this moral chiasmus.

Nietzsche leaves no doubt over which side was victorious in the historical struggle: the priests and slaves installed themselves as the universal class by imposing a group morality and a standardized set of (Christian) values over all (*GM*: I.7, 18). The means by which the priests achieved this revaluation was by introducing bad conscience, a turning inward of the nobles' domineering, affirmative outbursts that were originally aimed outward (*GM*: I.10, 20). Bad conscience universalized resentful, negating thinking by making it an interiorized principle of subjectivity. The moral chiasmus thus became operative within every individual: noble inclinations *as such* were evil and had to be turned inward. In this move of universalizing self-negation, the priests became the agents of what emerges as the *Genealogy's* true preoccupation in the third essay: the ascetic ideal. The ascetic ideal, for all its ambiguities (it can mean "nothing, or too many different things" with artists; *GM*: III.1, 68), shares several undeniable features with dialectical thinking: it seeks to overcome contradictions by deriving an affirmative moment from a negation (*GM*: III.7, 76-7); it aims for self-clarification by shedding illusions about itself and the world (*GM*: III.8, 80); it professes to enact a teleology toward a higher, utopian state (*GM*: III.11, 85); and it seeks to universalize itself (*GM*: III.23, 109). The

ascetic ideal as such is not equivalent to dialectics, but the priestly logic under the aegis of the ascetic ideal is the framework that would accommodate dialectics for Nietzsche. Nietzsche's own intervention consists in tracing the priests' strategic usage of chiasmic rhetoric, which, once exposed, divests dialectics of its veneer of logical authority and inevitability.

In unmasking the chiasmic trick that the priests played on logic again and again until they were firmly in power, the *Genealogy* takes a very different stance than Marx's *Critique*: instead of portraying history as dialectical, it reveals the historicity of dialectics. There is no logical structure to history that could be rendered in a propositional calculus, no Hegelian-Marxian "Vorstellung eines progressiven, emanzipatorischen Weltplanes" (Ruehl 56) that, ultimately, traps human agency in a debilitating teleology. Instead, rhetorical deployments of language create the appearance of an irrefutable logic that, once exposed, is neither conciliatory nor inexorable. Nietzsche's text embeds dialectics in a genealogical history and uncovers dialectics' reliance on the chiasmus as a rhetorical trick in its pursuit of an unattainable universality.

4. Vigor and Sickness: Thinking in a New Key

The contingent emergence and universalization of priestly thinking (and with it, a shallow Hegelianism in the terms of opposition and negation) notwithstanding, Nietzsche has to confront the urgent question: is there a way out? The outcome of the struggle between nobles and priests may not have been inevitable, but given that this universalization *has* occurred, is there anything in Nietzsche's writings that bears the potential to escape it? A regressive chiasmic reversal of the priests' logic with the aim to reinstate the noble one ("it is not true that the nobles are evil and, hence, that the priests and slaves are good; rather, the nobles are good and, hence, the priests and slaves are bad") is clearly not available to Nietzsche

since performing such a chiasmus would only constitute another operation of the priestly thinking he would like to escape. Instead, Nietzsche has to conceive of a regime of thinking that commands an entirely different logico-rhetorical apparatus, one which no longer seeks to present affirmation and negation as self-evident, clear-cut, and separable logical operators.

Indeed, such a regime is clandestinely present in the *Genealogy of Morality* from its very beginning. It seems to initially map onto the opposition of nobles and priests, while it is in fact wholly different and soon infiltrates the text with its own dynamic. This regime consists of the pair of vigor and sickness. Sickness brings with it pessimism, meaninglessness, an enfeebled will, passivity, and nihilism while vigor entails optimism, meaning, a strong will, and a propensity to act (*GM*: III.14). However, the two cannot be neatly separated in a polar opposition: vigor (which is not health) has a tendency to exhaust itself and degenerate into sickness. Once sickness has infused the healthy, there is no possibility of reversal (*GM*: III.14: 89). Vigor can only function as a temporary palliative (*GM*: III.17: 97). Sickness insidiously infects healthy thinking and operates unrecognized under the surface until it eventually comes to the fore.

Sickness and vigor thus do not operate on the same plane as noble affirmation or priestly negation. Even though Nietzsche tends to associate the former with the priests (cf. 16, 92) and the latter with the nobles (*GM*: I.4, 14), sickness and vigor create their own unique and independent dynamic that infuses the other two logics and runs roughshod over them. Nietzsche repeatedly emphasizes that there were times when the priests, bad conscience, and the ascetic ideal infused life, promoted vigor, and provided meaning:

In fact, the old depression, heaviness and fatigue were thoroughly *overcome* by this system of procedures, life became *very* interesting again: awake, eternally awake, ... already people were no longer making complaints *against* pain, they *thirsted* for it; “*more* pain! *more* pain” screamed the desire of his disciples and initiates for centuries (*GM*: III.20, 105; cf. also II.18, 59-60).

By the same token, it was the nobles' flagging of vigor that precipitated their defeat at the hands of the priests (*GM*: III.16, 94-5). Priestly thinking entered the scene as a sickness to noble thinking, while its own sickness had not yet broken out. Nietzsche wrestles with this ambiguity in III.15:

[The priest] must be sick himself, he must really be a close relative of the sick and the destitute in order to understand them, – in order to come to an understanding with them; but he has to be strong, too, more master of himself than of others. (92)

Positive affirmation and oppositional negation can thus *each* occur in the mode of vigorous or sick (irrespective of the fact that one of them has the capacity to increase sickness, while the other one does not). Both logical apparatuses face the dilemma that their vigor tends to flag over time. Nietzsche calls this danger of losing steam the fundamental condition of life:

The more normal this sickliness is in [humanity] – and we cannot dispute this normality –, the higher we should esteem the unusual cases of spiritual and physical powerfulness, [humanity's] *strokes of luck*, and the better we should protect the successful from the worst kind of air, that of the sickroom. (*GM*: III.14, 89)

Vigor can only ever be a temporary remedy for an insidious operation. As per its own historicity, vigor runs out and what remains — is sickness.¹¹

Nietzsche's history of the struggle between two particular logics thus unfolds within humanity's much more fundamental (and often futile) struggle to resist sickly thinking. While the first essay starts from the nobles' perspective and the second from the priests and slaves', the third essay stages sickness' rise to prominence: words with the stem "krank" occur a total

¹¹ It may seem that, at first glance, the advent of sickness, bringing about the defeat of the nobles at the hands of the priests, institutes a dialectical transition from the noble to the priestly logic. However, sickness is thoroughly a-dialectical: the transition from the former to the latter is not inexorable, but could have been avoided; sickness does not resolve any contradictions but exacerbates them; and it does not lead to a permanent state of sublation on some higher historical plane. Sickness brings about confusion, not progress.

of 24 times in the preface and the first two essays, while they appear 75 times in the third.¹² At bottom, Nietzsche sought to grasp the debilitating nihilism of his age that arose from the sickness of priestly Hegelianism, a heuristic move “reflecting the wider biologisation and medicalisation of nineteenth-century cultural discourse” (Moore 193).

If sickness and vigor, not dialectics, are the proper conduits of history for Nietzsche, and if he seeks to stage his own historical intervention by providing an antidote (or palliative) to the crippling sickness of historical epigonism and nihilism, the *Genealogy* should not merely trace the advent of sickness, but aim to infuse vigor into thinking, to establish a “counter-memory—a transformation of history into a totally different form of time” (Foucault 1977: 160), one in which events are not discovered, but have an effect on us (Foucault 1977: 154). Indeed, Nietzsche indicates in his preface that invigoration is the underlying objective of his treatise. In response to Paul Rée’s study on the origins of morality, in which he detects a certain “pessimism and fatigue” (*GM: P 7, 8*), Nietzsche wants to examine problems of morality in a way to “one day [be] allowed to take them cheerfully” (*GM: P 7, 8-9*). With this attitude, he wants to jolt our thinking so to enable us to see things differently:

The vast, distant and hidden land of morality – of morality as it really existed and was really lived – has to be journeyed through with quite new questions and as it were with new eyes: and surely that means virtually *discovering* this land for the first time? ... (*GM: P 7, 8*; cf. also II.16, 57/8 and III.19, 103)

Nietzsche closes his preface with the exhortation to listen: “– If anyone finds this script incomprehensible and hard on the ears, I do not think the fault necessarily lies with me” (*GM: P 8, 9*).

¹² Figures obtained from a count on [nietzschesource.org](http://www.nietzschesource.org/#eKGWB/GM): <http://www.nietzschesource.org/#eKGWB/GM>. Last accessed March 7, 2021.

Reading the *Genealogy* thus becomes an exercise in listening to the text's surging and flagging of energy beyond the propositional content of its arguments. It is the *Genealogy's* rhetorical arsenal (beyond the schematics of the chiasmus) that has to be interrogated for its ability "to advance in the manner of a spur of sorts (*éperon*)" (Derrida 39) in pursuit of a constantly eluding truth, and its blurring of the boundaries between what it states and what it performs (de Man 1979b, 131). It is on this risky endeavor of philosophizing in a new key that he stakes the success or failure of breaking out of the languishing sickness of Hegelian nihilism. We will thus trace Nietzsche's *performance* of how the noble and priestly logics vigorously surged and eventually degenerated into nihilism, before examining the key rhetorical vehicle with which he wants to escape this nihilism: aposiopesis.

The noble logic asserts itself and indirectly points to a dimly defined opponent. Its initial state is strength and joyful vigor. Indeed, Nietzsche himself declaims in his preface: "— what business is it of mine to refute! —" (*GM: P* 4, 6). Unsurprisingly, the first essay opens with a self-assured mocking of the English moral psychologists: "These English psychologists [...] have a significant advantage over their books — *they are actually interesting!* These English psychologists — just what do they want?" (*GM: I.1*, 10). Instead of seriously attempting to refute their arguments, he ridicules their predilection for studying taboo topics while living in a prudish society (*GM: I.1*, 10) and their amateurish methodology:

So you have to respect the good spirits which preside in these historians of morality! But it is unfortunately a fact that *historical spirit* itself is lacking in them, they have been left in the lurch by all the good spirits [von all guten Geistern verlassen: abandoned by all good and common sense — TK] of history itself! (*GM: I.2*, 11)

Nietzsche thus summarily dismisses "[t]he idiocy of their moral genealogy" (*GM: I.2*, 11). He speaks with the joyful disdain of someone who knows of his own superiority, which allows him to vigorously affirm his own conviction about the origin of the values of good and bad:

Now for me, it is obvious that the real breeding-ground for the concept ‘good’ has been sought and located in the wrong place by this theory: the judgment ‘good’ does *not* emanate from those to whom goodness is shown! (GM: I.2, 11, first emphasis mine)

This line of thinking continues with a choppy but confident philological speculation on the origins of the words good and bad (GM: I.4): the nobles affirmed themselves as good in their violent contests with the slave caste, who thereby received the label “bad.” Nietzsche discusses the vigorously effusive, disdainful, and petulant nobles in a tone that is itself vigorously effusive, disdainful, and petulant.

Priestly thinking is reactive, spiteful, keenly aware of its opponent, and arises insidiously after having been in hiding. As such, it makes its appearance insidiously in Nietzsche’s text. Without having fully developed the account of the nobles yet, Nietzsche shifts his attention to the priests when he suggests that their assertion of its mode of valuation “does not yet constitute an exception to the rule” (GM: I.6, 15), but soon develops into an oppositional pair with the nobles (GM: I.7, 16-7). Having heard about the nobles first, the reader is initially led to believe that they lived without any internal threat, but now learns that the priests have been conspiring among the nobles from the very beginning.

With this shift of focus, Nietzsche’s own manner of arguing turns more contentious and oppositional as well. He fleshes out the priestly logic by contrasting it with the noble one (GM: I.10, I.11, I.16). Unlike his mocking dismissal of English moral psychologists, his condemnation of the priests is venomous and resentful, implying that, because the priests are inferior, the nobles are superior. He resentfully laments that, under the priests, “[the human being]’ is first and foremost a teeming mass of worms; that the ‘tame [human]’, who is incurably mediocre and unedifying, has already learnt to view himself as the aim and pinnacle, the meaning of history, the ‘higher [human]’ (GM: I.11, 24). After priestly thinking has crept into and infected Nietzsche’s noble way of thinking in a moment of distraction, the joyful

levity of the first few pages disappears, as he now argues about the priests in a priestly manner. While his own energy is still high at this point (just as the priests were initially vigorous – cf. *GM*: III.20), it is now only a matter of time for it to flag.

A first sign of losing his rhetorical stride appears in I.12, which begins with the confession that, “[a]t this juncture I cannot suppress a sigh and one last hope” (25) but ends with the diagnosis that humanity is sick of itself: “The sight of [the human] now makes us tired – what is nihilism today if it is not *that*?... We are tired of [*the human*] ...” (25). Nietzsche is unsure whether there can be another incarnation of the noble after the historical appearance of Napoleon, “this synthesis of *Unmensch* (brute) and *Übermensch* (overman)” (*GM*: I.16, 33, German in the translated version), but underscores the possibility of “an even more terrible flaring up of the old flame” between the noble and the priestly (*GM*: I.17, 33). The vigorous mode is still palpable in the beginning of the first essay, but starts to give in to uncertainty and disorientation. The first essay does not yet extensively discuss sickness, but hints at its advent toward its end.

Sickness appears more prominently in Nietzsche’s diction after a turning point in the second essay: the appearance of bad conscience, the “serious illness to which [humanity] was forced to succumb by the pressure of the most fundamental of all changes which he experienced” (*GM*: II.16, 56).¹³ Discussing this paradoxical notion of domineering forces turned inward, Nietzsche vacillates between the prospect of a “second innocence” (*GM*: II.20, 62) and a fear of falling into complete and irreversible nihilism (*GM*: II.21, 63), which causes him to condemn bad conscience as “the most terrible sickness ever to rage in [humanity]” (*GM*: II.22, 64). His uncertainty reaches a temporary apex when he closes the second essay

¹³ This passage is located almost at the center of the *Genealogy*. In the German original, words with the stem “krank” occur 16 times before it, and 83 times after. Figures obtained from a count on [nietzschesource.org](http://www.nietzschesource.org/#eKGWB/GM): <http://www.nietzschesource.org/#eKGWB/GM>. Last accessed March 3, 2021.

with “three question marks” (*GM*: II.24, 65) and muses on life under the aegis of bad conscience: “A reverse experiment [of reinstating the noble logic – TK] should be possible in *principle* – but who has sufficient strength? –” (*GM*: II.24, 66). The second essay ends on the desperate cry that the *Übermensch*, “*he must come one day ...*” (*GM*: II.24, 67), but immediately lapses into resignation: “– But what am I saying? Enough! Enough!” (*GM*: II.25, 67). Nietzsche’s discussion of sickness is itself sick: among desperate flashes of energy, his vigor begins to decline.

The drama of surge and decline is most openly played out in the third essay, which comes fully under the aegis of sickness. In discussing the ascetic ideal in its different manifestations for artists, philosophers, women, weak average people, priests, and saints, the same pattern is performed again and again: an erstwhile spike of energy is followed by an appearance of self-contradiction that saps the initial vigor. While Nietzsche is rather glib about this phenomenon at first (“What do ascetic ideals mean? – With artists, nothing, or too many different things”; *GM*: III.1, 68), he gradually renders each iteration with less energy, leading him to muse: “– *life itself must have an interest* in preserving such a self-contradictory type. For an ascetic life is a self-contradiction” (*GM*: III.11, 86). The priests initially promoted life (*GM*: III.13, 88), but ultimately made “the sick person even sicker” (*GM*: III.20, 104). Nietzsche performatively stages how, among flashes of energy, the ascetic priestly logic exhausts itself in constantly turning against itself.

Nietzsche thus homes in on the debilitating paradox of self-defeat. By III.20, he begins to suspect that he, the genealogist of priestly thinking, is still beholden to the ascetic ideal himself: “probably we, too, are still the victims, the prey, the sick of this contemporary taste for moralization” (103). Nietzsche now enters truly dangerous territory, as he exposes science and the search for truth, the alleged counter-ideals to the ascetic ideal, as its secret accomplices:

it “permits of no other interpretation, no other goal, and rejects, denies, affirms, confirms only with reference to *its* interpretation” (GM: III.23, 109); “Where is the *counterpart* to this closed system of will, goal and interpretation? Why is the counterpart *lacking*? ...” (GM: III.23, 109). Nietzsche runs up against the paradox that there seems to be no opposite to oppositional thinking.

The priests’ incessant urge to negate under the ascetic ideal in search for truth has left nothing in place except the ascetic ideal itself, i.e. the negating will to truth. This will now faces the most extreme implication of its pursuits: its self-negation. But whether that entails an extreme invigoration (willing the nothing) or a final desiccation (not willing anymore) is unclear. In III.27, Nietzsche indicates that the will to truth is on the brink of overcoming itself:

All great things bring about their own demise through an act of self-sublimation: that is the law of life, the law of *necessary* ‘self-overcoming’ in the essence of life [...] we stand on the threshold of *this* occurrence [...] from now on, morality will be *destroyed* by the will to truth’s becoming-conscious-of-itself: [...] the most terrible, most questionable drama but perhaps also the one most rich in hope... (119)

Yet, in the final sentence of the penultimate section, Nietzsche seems to retract his optimistic assertion that the negating will to truth can negate itself, and confines it to willing nothingness instead: “And, to conclude by saying what I said at the beginning: [humanity] still prefers to *will nothingness*, than *not* will...” (GM: III.28, 120). Nietzsche has drawn out the dilemma in its most explicit form: does the will to negate ultimately negate itself such that there is no more willing, or does it end up in willing a reified nothingness? Does dialectics overcome itself (and thus cease to be dialectical), or does its alleged self-overcoming constitute the performance of another step within the dialectic? Nietzsche does not decide between the two positions. The chiasmic structure of dialectics (reversing two terms and universalizing one while eliminating the other) blurs into indeterminacy once it is applied to itself.

5. Gesturing into (the) Nothing: Aposiopesis and the Rhetoric of Silence

While Marx sought to enlist a rhetorical device in the service of logic, Nietzsche fully embraces the primacy of rhetoric, recognizing that whichever way of thinking takes on the self-evidence *of* a logic depends on the rhetorical matrix that determines the presuppositions for what is persuasive in the first place. The thinking that appears to be logical is the one that can be espoused with the vigor of self-evidence, while a logic becomes unpersuasive once it has exhausted itself and turned sick. Returning to Breazeale's question, affirmation and negation thus acquire wholly different senses under Nietzsche: affirmation in the new sense is the effect that is created by a logic that has the rhetorical verve to argue its claims vigorously, be it a noble logic (privileging affirmation in the old sense) or a priestly one (privileging negation in the old sense). Negation in the new sense occurs when a logic lacks the vigor to set a rhetorical context in which such an effect of affirmation could be produced. The *Genealogy* thus undertakes to make its readers experience (indeed, inhabit) different logics in the modes of vigorous and sick.

In the course of this performance, Nietzsche has staged the exhaustion of the noble and, especially, the priestly logic, leaving it unclear whether he is able to recuperate a vigorous attitude in the face of a universal sickness of thought: Nietzsche may cease to write after III.28 because he is exhausted and does not know how to go on, or because he is making an ominous gesture into the undiscovered lands of morality that he mentioned in his preface. III.27 and III.28 both end in portentous ellipses "...", without a clear indication of what they are hinting at: the *Genealogy* ends by gesturing into (the) nothing.

Nietzsche's penultimate falling silent puts *us* readers to the test—will we despair, or be invigorated? Will we see no way out, or multifarious new openings? As we waver between nihilism and vigor in those moments, the path we ultimately follow will reveal if we belong to

the sick or the vigorous, the noble or the priestly. The reader thus has to press even beyond Nietzsche's tone and style and into the silent interstices of his philosophy. Nietzsche circles around a problem, "which can justifiably be called a *quiet* problem and fastidiously addresses itself to only a few ears" (*GM*: I.5, 13). Consider his rebuke of the careless inference that priestly-slavish thinking has won:

This is the epilogue by a 'free-thinker' to my speech [...]; [s/he] had clearly listened to me up to that point, and could not stand listening to my silence. As a matter of fact, there is much for me to keep silent about at this point. – (*GM*: I.9, 19)

It is in those moments of silence, that we have to listen without Nietzsche's guidance. Out of the 78 sections in *The Genealogy*, 37 end in ellipses. 147 further ellipses are distributed throughout the text.¹⁴ Nietzsche's pauses and ellipses are nefarious because they do not merely indicate that *something* has been left unsaid, that a thought could be filled in by grasping the logic and thrust of what has been said — more radically, they may falsely indicate that there is *anything* to be thought *at all*. An ellipsis could be a sudden, embarrassed realization and silent admission that *nothing* can be said or thought anymore. While Marx's stock figure of the chiasmus made explicit what it anticipated, Nietzsche constantly draws on *aposiopesis*: falling silent and leaving it perfectly open whether anything is anticipated at all, or if thought has simply collapsed. Aposiopesis is the dangerous rhetorical device that tests our courage if we can peek into the abyss, if we can move on without handrails.

Consider the three ellipses and their ambiguous possible connotations in this passage:

more comfortable, more mediocre, more indifferent, more Chinese, more Christian – no doubt about it, [the human being] is getting 'better' all the time ... Right here is where the destiny of Europe lies – in losing our fear of [the human being] we have also lost our love for [it], our respect for [it], our hope in [it] and even our will to be [human]. The sight of [the human being] makes us tired – what is nihilism today if it is not *that*? ... We are tired of [*humanity*] ... (*GM*: I.12, 25)

¹⁴ Figures obtained from a count on [nietzschesource.org](http://www.nietzschesource.org/#eKGWB/GM): <http://www.nietzschesource.org/#eKGWB/GM>. Last accessed March 8, 2021.

The first ellipsis comes after a cascade of negatively loaded epithets, all of which refer to unworthiness, docility, and lack of energy for Nietzsche. Nietzsche himself seems to tire as he makes the enumeration. Yet, he ends it in an ironic jab at a “constantly improving” humanity: the first ellipsis still carries undertones of noble disdain. His energy surges again when he sounds the alarm bells about Europe’s destiny, trying to alert the reader that tiring of humanity is a great danger. The second ellipsis comes after an emphasized “that” and a question mark, strongly gesturing that there is something momentous to grasp and, more importantly, to resist. Yet, this energy flags again in his assertion that “[w]e are tired of [*humanity*] ...,” which sounds resigned and without direction, not gesturing at anything. Nietzsche’s aposiopeses variously show remnants of a noble attitude, an energetic oppositionality, and a sickly resignation.

If an ellipsis is not merely ambiguous as to what it points to, but is ambiguous as to whether it points to anything at all, what we hear in the ellipsis will not so much be what Nietzsche wanted us to hear, but rather reveal who we are — to others and to ourselves. What we hear Nietzsche say in those silences depends on how well we have listened when he spoke. If we have understood, indeed *experienced* that there are other ways of thinking than the oppositional and dialectical, the implosion of the chiasmus and the advent of paradox do not deter us. We joyfully posit our own categories. If we can only conceive of Nietzsche’s distinctions as antagonistic oppositions, we are consigned to priestly nihilism, not knowing how to go on. Nietzsche is the thinker of danger: lively nobles will rejoice at him; lively priests will be reinvigorated in opposition to him; the sick will become sicker with him.

While the point of Marx’s chiasmi is to guide us along a path he saw as predetermined by history, Nietzsche’s aposiopeses test us if we believe that there can be an *open* future and a history that “refuses the certainty of absolutes [and] is capable of liberating divergence”

(Foucault 1977: 153). The question of whether Nietzsche is a dialectical thinker obscures the multifarious modes and registers in which he philosophizes. Nietzsche tempts himself – and us – with the danger *of* dialectical thinking, leaving it up to us to find our way out again. If Nietzsche’s writings succeed in expanding our notion of what it means to philosophize to include the rhetorical matrix that conditions the self-evidence of a particular logic, and if we are able to understand affirmation and negation from their sources (vigor and sickness) that ground any rhetorical posture and the logical operations that this posture permits, then Nietzsche becomes much more than a dialectical (or anti-dialectical) thinker to us. Indeed, his thinking begins to harbor the promise of moving exactly beyond such a stale oppositionality. If, on the other hand, Nietzsche’s rhetoric fails to exert a hold over us, we remain trapped in the machinery of dialectics.¹⁵ Where Nietzsche takes us is a matter of what we hear in...

PART II: LIVING IN THE ABSENCE OF SYSTEMS: NIETZSCHE’S ETHICS OF DANGER¹⁶

1. Introduction

In September 1886, the poet and critic Josef Victor Widmann penned a review of *BGE* entitled, “Nietzsche’s Dangerous Book” (Widmann 618-23), observing that a keen sense of danger pervaded the work (Widmann 620). Nietzsche, who often felt misunderstood and wrongfully attacked, responded enthusiastically to Widmann on June 28, 1887:

Mainly, I have to thankfully inform you, after a year’s time no less, that your review has been by far the most “intelligent” one that this uncongenial book has received

¹⁵ As such, Derrida’s claim that Nietzsche’s styles function as spurs to our thinking is taking it for granted that his rhetoric succeeds in spurring us in *some* direction. Depending on the reader, however, such a success might or might not materialize. Similarly, with regard to Foucault’s claim on the historically transformative movement of Nietzsche’s genealogy, it becomes clear that such a transformation may or may not occur with any given reader. Nietzsche takes a risk: he exposes himself to danger.

¹⁶ This part was originally published by the Penn State University Press as: Kuchne, Tobias. “Nietzsche and the Ethics of Danger.” *Journal of Nietzsche Studies* 49.1 (2018). 78-101. Print. Reprinted with permission of the publisher.

until now. Poets are, after all, “divinatory” beings: in the end, such a book of riddles will much sooner be deciphered and “cracked open” by a poet, rather than by a philosopher and “specialist.” *KSB* III 5:869, 101-02 (my translation).

Indeed, Widmann identified a motif that runs not only through Nietzsche’s *BGE*, but throughout his writings as a whole.¹⁷

Despite its ubiquity, however, the theme of danger has largely been overlooked in the Nietzsche scholarship. Ottmann’s *Nietzsche-Handbuch* does not list “Gefahr” as an entry, and Niemeyer’s *Nietzsche-Lexikon* features only a short discussion of the fascist exploitation of his slogan, “Live dangerously!” (Frischmann 101). Very few studies have even touched on danger and its role in Nietzsche’s philosophy. Conway’s *Nietzsche’s Dangerous Game* traces how Nietzsche’s post-1885 writings exhibit the very decadence that he lamented of his time (1-6). Müller-Lauter’s study on Nietzsche’s philosophy of oppositions mentions the dangers that Nietzsche sees in a debilitating historical sense, self-knowledge, and herd morality’s hostility and mistrust of non-conformers, but it does not expand upon those observations (39-40, 69, 95). And Joisten argues that, in Zarathustra’s account of the human as a self-overcoming being, danger crucially figures into Nietzsche’s attempt to move beyond anthropocentrism (70-97). While those studies gloss danger’s unpredictable nature and humanity’s thorough exposure to it, none of them recognize its centrality in Nietzsche’s philosophy. This calls for a closer study of Nietzsche’s notion of danger, not only because of his own emphasis on it, but because his thinking has posed dangers to commentators on two fronts: first, his paradoxical reasoning, and second, his unsettling ethics.

¹⁷ A search for *Gefahr* and its derivatives, *(un)gefährlich*, *(un)gefährdet*, etc., on <http://www.nietzschesource.org> (accessed March 4, 2021) in a few representative works of Nietzsche’s yields, for example, 20 occurrences in *BT*, 10 in *DS*, 28 in *HL*, 48 in *SE*, 80 in *HH* II, 68 in *GS*, 101 in *BGE*, and 37 in *GM*.

This part treats the motif of danger as the fulcrum between Nietzsche's mode of reasoning and his ethics. Danger is the threat of unexpected decline, dissolution, or the destruction of a previously stable structure. It is the pervasive, ineliminable menace that any reassuring worldview—such as metaphysical systems or dogmatic moralities—may turn out to be irredeemably paradoxical, i.e., self-contradictory and indefensible, on its own terms. From the earliest works onward, Nietzsche seeks to alert his readers to this danger and its vast extent. In *BT*, *DS*, and *HL*, danger emerges as a ubiquitous, but self-effacing condition that becomes more menacing the longer it goes unnoticed. Danger derives its potency from the fact that its pervasiveness is underestimated. As Nietzsche's thinking evolves, he uncovers the ethical implications of our dangerous condition. In *SE*, he begins to explain our task of confronting danger as an ethical one: we are to live our lives courageously and not shrink from the abyss of paradox.¹⁸

In his middle works *HH*, *D*, and *GS* I–IV, Nietzsche gives substance to this imperative by confronting this ethical task through the idea of the free spirit, adopting his ethos of courageous intellectual honesty (*Redlichkeit*), and exposing himself to danger. Indeed, far from banishing danger, it is only in this dangerous space that a *genuine* ethics can even emerge for him. Nietzsche's philosophy is a *philosophy of danger*—it advances the dangerous proposition that an ethics (in a novel sense) can arise only in moments when metaphysics collapses. He puts into action the dangerous doctrine that intellectual honesty drives us to flirt with paradox and reveal the self-contradictions of any metaphysically grounded morality, a pursuit that does not dispel the ethical question, but *raises* it with unprecedented urgency: in absence of all

¹⁸ This chapter therefore builds on Müller-Lauter's claim that Nietzsche's thinking was prone to encounter tensions and contradictions whenever he had seemingly reconciled them (Müller-Lauter 7): Nietzsche's critical urge often drew him toward those debilitating paradoxes in order to confront them.

security, what am *I* to do? Only when we are endangered is it possible to hear the call to a genuine ethics.

2. Establishing the Task of Confronting Danger: Nietzsche's Early Works

In his 1886 preface to *BT* (1872), Nietzsche disavows much of his youthful work. Nonetheless, he stresses that he had already “got hold of [. . .] something fearsome and dangerous [. . .] a *new* problem: [. . .] the *problem of science* [*Wissenschaft*] itself” (*BT* P:2). This problem crystallized for Nietzsche in tragedy’s degeneration into an overly rationalist, self-aware staging of its own form, which grew out of the complex, asymmetrical relation between the Apollonian and the Dionysian. The Dionysian force of obliteration, excess, and turbulent presence undergirds the Apollonian principle of individuation, measure, and lucid appearance. For Nietzsche, Apollo was a protective figure for the Greeks against the “dangerous power” of the “crude, grotesque manifestation of the Dionysiac” (*BT* 2). Apollonian art forms provide a relief in moments when the Dionysian assails the willpower that is necessary for an ordered, purposeful life: “[h]ere, at the moment of supreme danger for the will, *art* approaches as a saving sorceress with the power to heal. Art alone can re-direct those repulsive thoughts about the terrible or absurd nature of existence into representations with which man can live” (*BT* 7). However, those clear Apollonian forms emerge only as a fleeting “metaphysical solace” from troubled the Dionysian (*BT* P:7). The permanence of those artistic forms is an illusion, albeit a necessary one that gives meaning to human existence (*BT* 5). The Apollonian banishes the danger of the Dionysian, but for a limited time.

The tragic hero insists that he is an autonomous, lucid individual who rises above the Dionysian chaos. He is tragically unaware that he has emerged out of the Dionysian and is “merely [a] mask of that original hero, Dionysos” (*BT* 10). The tragic hero is constantly

threatened by Dionysian obliteration, and his quest for self-knowledge must end in the self-destructive insight that the very assumption on which he has based his identity—that he is an autonomous agent in control of his fate—is illusory: his self-fulfillment coincides with his self-annihilation, which renders his existence tragic. The tragic hero embodies a dilemma: if he ceases to insist on his autonomy, he loses the struggle against Dionysian obliteration immediately, yet persisting in his quest can only lead to the inevitable realization that there is no escape from the Dionysian. The hero survives only as long as he can withstand the existential danger of the Dionysian and delay its inevitable victory.

According to Nietzsche, it is not just the tragic hero who blindly rushes into self-destruction, but also Greek tragedy itself. With Euripides' plays, tragedy began to consciously reflect its structural pattern of a hero confronting, misunderstanding, and eventually succumbing to danger. By rationalizing the hero's struggle against the Dionysian on the stage, "[t]he *Apolline* tendency has disguised itself as logical schematism" (BT 14). Tragedy considered itself to have achieved a permanent victory in the name of the Apollonian over the Dionysian, repeating the same hubristic error as the tragic hero himself. However, just as with the tragic hero, it is thus only a matter of time until this Socratic Apollonianism, with its hypertrophied logic, recognizes *its* self-deception and "sees how logic [. . .] finally bites its own tail" (BT 15). In Nietzsche's words: "Greek tragedy [. . .] died by suicide, as the result of an irresolvable conflict, which is to say tragically" (BT 11). Tragedy, even though it enacted the hero's dangerous path again and again, was unable to grasp danger in such a way that it could banish it. In fact, tragedy as an art form succumbed to danger in the very moment it thought to have mastered it in a lucid conceptual structure. The paradoxical tendency is the new "fearsome and dangerous" problem that Nietzsche recognized: the danger of self-destruction strikes exactly when it is *least* expected (BT P:2).

Nietzsche tackles danger, its elusive nature, and the problem of how to recognize and confront it more explicitly in the *Untimely Meditations*. *DS* (1873) opens with a warning to the German people about their recent victory over France: “Public opinion in Germany [. . .] notwithstanding, it has to be said that a great victory is a great danger” (*DS* 1, 3). The “great danger” consists in the German people’s mistaken assumption that their military victory was due to their cultural superiority when, in fact, their culture is in dire straits. Nietzsche fears that Germany’s military victory has lulled them into a false sense of security, risking to turn “our victory into a defeat: *into the defeat, if not the extirpation, of the German spirit*” (*DS* 1, 3). The explicit goal of *DS* is to alert the German people to the danger embodied by the alleged guardians of its culture: the cultural philistine. The cultural philistine is a shallow thinker who is oblivious to contradictions and inconsistencies in his own conduct and belief systems (*DS* 2, 8). He opposes genuine curiosity and creativity (*DS* 2, 9–10), while unquestioningly adopts anything from the past (*DS* 2, 11). The philistine promotes “the opposite of culture, barbarism, [which is a] lack of style or a chaotic jumble of all styles” (*DS* 1, 6). Culture endangers itself by entrusting its protection to agents who secretly undermine it: its demise lurks where it is least expected.

Direct attacks from such defenders of true culture as Nietzsche have no effect on the cultural philistine. When called on the carpet, he is offended (*DS* 10, 45) and negates any challenge to his incoherent thinking, thereby continuing to promote cacophony and barbarism. Yet Nietzsche attacks the widely respected David Strauss. Nietzsche fiercely denies that Strauss is a rigorous thinker (*DS* 9, 44) or a skillful writer (*DS* 10, 48), and he frequently resorts to invective, especially in the penultimate section’s disproportionate and repetitive condemnation of Strauss’s stylistic and grammatical gaffes. Instead of putting forth a stylistically coherent example of culture, Nietzsche’s diatribe displays precisely the barbaric

shortcomings he criticizes in Strauss: negating, petty, self-righteous, logically inconsistent, and stylistically wanting, while its author is convinced that he is a champion of culture. Nietzsche, just like Strauss, seems to endanger culture.

Toward the end, however, Nietzsche indicates that he has not fallen into the trap of philistinism: “But he who absolutely refuses to heed this warning, and insists on continuing in his belief that Strauss is a classic, should [...] imitate him. If you try to do this, though, it will be at your peril [auf eigene Gefahr]” (*DS* 11, 53). If the philistine uncritically imitates the past, then Nietzsche “out-philistines” the philistine with his conscious, parodying, and critical imitation of supposedly cultured public discourse. Nietzsche one-ups the dangerous philistine by donning his mask, drawing attention to it, and thereby exposing it. For those who can see, Nietzsche brings the inherent, dangerous paradoxicality of cultural philistinism out into the open.

DS thus clarifies the problematics of danger and suggests a possible way of addressing it. As embodied by the cultural philistine (and painfully experienced by the tragic hero), danger hides where one least expects it. In the moment of a perceived cultural victory, the philistine has already eroded culture. Danger derives its potency precisely from the fact that it is so difficult to locate. It requires a vigilant spirit that constantly looks for inconsistencies and paradoxes. However, simply identifying a danger does not dispel it, as it can efface itself again and withdraw elsewhere. Rather, exposing danger requires imitating the adversary to the point of the grotesque for others to see him as the menace he is. While this risks becoming indistinguishable from the adversary, there is no other choice: confronting danger is a dangerous endeavor.

HL (1874) expands into a more general treatment of how to awaken a people to the unrecognized danger that surrounds them. Of each sense by which one can relate to history—

the monumental, the antiquarian, and the critical—there exists a life-serving form that fosters vigor and a keen awareness of one’s historical moment, and a degenerate form that stifles life, endangers it, and hastens its demise. Moreover, each historical sense’s preservative forms already harbor within themselves the danger of collapsing into their opposites.

The life-serving monumental sense inspires grand deeds by providing examples of past achievements, while its degenerate counterpart cuts the past off from the present and thus “incurs the danger of becoming somewhat distorted, beautified and coming close to free poetic invention” (*HL* 2, 70). Past achievements devolve into mere mythical fictions belonging to a bygone era of deeds that can no longer be emulated (*HL* 2, 71). The life-serving antiquarian sense can counteract such a loss by maintaining a vital connection with the past, but threatens to turn against life in a different way. “This always produces one very imminent danger: everything old and past [is] taken to be equally worthy of reverence, while [...] everything new and evolving, is rejected and persecuted” (*HL* 3, 74). The antidote to such a mummification of life (*HL* 3, 75) is the life-serving critical sense, whose negation of the past can liberate life by exploding the unquestioning antiquarian eclecticism (*HL* 3, 75–76). However, the criticism’s transgressive impulse can easily become excessive: “It is always a dangerous process, especially for life itself: and men and ages which serve life by judging and destroying a past are always dangerous and endangered men and ages” (*HL* 3, 76). A critical experiment is “always a dangerous attempt because it is so hard to know the limit to denial of the past” (*ibid.*). Instead of atrophying life, the degenerated critical sense might destroy it altogether.

In his current historical situation of antiquarian degeneration, however, Nietzsche sees no choice but to harness the critical historical sense for life’s reinvigoration. He thus devotes more than half of *HL* to this therapeutic goal. Section 5 begins: “The oversaturation of an age

with history seems to me to be hostile and dangerous to life in five respects” (83). Those five dangers are: a weakened personality, a false sense of historical objectivity, modern man not coming into his own, the belief of belonging to a generation of disempowered epigones, and slipping into a “dangerous mood of irony in regard to itself and subsequently into the even more dangerous mood of cynicism” (*ibid.*). Nietzsche addresses each danger in a separate section by rhetorically exposing their paradoxical, self-undermining structure.

While the strategy is similar in each attempt, the fifth danger receives the most interesting treatment. Having trained his readers in critically identifying the dangerous, self-destructive tendencies of the degenerate antiquarian sense, Nietzsche hazards a more allusive approach in section 9, requiring the reader to undertake the necessary critical work on her own. The fifth danger to which life has succumbed under antiquarian history is a sense of irony, which Nietzsche defines as the fear of being unable to preserve one’s youthful hopes, and of cynicism, i.e., the conviction that the present could not have turned out differently than it has (*HL* 9, 107–8). In other words, degenerate antiquarian history devalues the present by removing historical agency over past and future events.

Just as in the first *UM*, Nietzsche dangerously parodies his adversary. He resorts to irony and cynicism by laconically deriding the philosophy of Edouard von Hartmann, “in whose head the age has come to an ironical awareness of itself” (*HL* 9, 108). He ridicules Hartmann’s cynical theory that history inexorably runs along a predetermined course, and that now, close to its end, the only option is to resign oneself to history’s necessity (*HL* 9, 109, 110). Nietzsche retorts: “I am quite willing to vote with the majority [...] that next Saturday night punctually at twelve o’clock the world shall perish [...] from tomorrow there shall be no more time” (*HL* 9, 112). Nietzsche’s quip, which amounts to asking, “What happens after time has ended?” draws out the ridiculous inconsistency of a stifling past bringing history to a

halt. Nietzsche takes the dangerous self-contradictory structure of antiquarianism—history stymying history—and critically turns it back on itself.

Nietzsche toys with dangers in order to overcome them, as he claims in the last section:

Land! Land! Enough and more than enough of the wild and erring voyage over strange dark seas! At last a coast appears in sight: we must land on it whatever it may be like [...] This voyage was perilous [gefährlich] and exciting. [...] In pursuit of the perils [Gefahren] of history we have found ourselves most acutely exposed to [them]. (*HL* 10, 116)

Nietzsche here associates perils (*Gefahren*) with the metaphor of a ship voyaging (*fabren*) into unknown waters. While he considers himself to have reached safe, solid ground again, he concedes: “I have no wish to conceal from myself that, in the immoderation of its criticism, [. . .] in its frequent transitions from irony to cynicism, [. . .] the present treatise itself reveals its modern character, a character marked by weakness of personality” (*ibid.*). Alleviating the degeneration of the antiquarian sense runs the risk of falling into an excessive, destructive critical sense. While *HL* continues the strategy of confronting danger first explored in *DS*, it is aware of the historical subject’s dilemma: giving free rein to the contemporary historical sense (ignoring danger) guarantees eventual degeneration, while promoting a different historical sense (confronting danger) risks excess. Addressing one danger entails exposing oneself to another, revealing that danger is inscribed into history itself and can never be permanently dispelled: the only choice we have is to ignore or confront it.

Nietzsche’s early philosophy is a philosophy of danger. Tragedy, culture, and history are beset by a hidden, innate paradoxical tendency towards self-destruction. Ubiquitous but clandestine, danger lurks where it is least suspected, making it impossible to confront it from a position of security. Facing it is an inherently dangerous endeavor that follows no rote formula. While *BT* and *DS* engage specific historical phenomena, *HL* expands danger’s scope to history as such. (Put differently, whether one is a post-classical Greek or a German citizen

enjoying the political and cultural successes of the early 1870s, danger lurks in *any* historical situation.) One has to be vigilant at *all* times to discern where the threat of degeneration may be hiding.

In *SE* (1874), Nietzsche states danger's ubiquity in even clearer terms: "to live at all means to live in danger" (*SE* 3, 144). Danger does not pervade a particular historical period, or history in general, but life as such. Nietzsche approaches this problem by developing three human types—the Rousseauian, the Goethean, and the Schopenhauerian—each of which risks falling short of achieving its full potential by succumbing to its peculiar danger. The emergence of danger as a fundamental condition of all aspects of life induces Nietzsche to confront the *ethical* question: if to live means to be in constant danger, how is one to act and live under it?

Nietzsche undertakes a first attempt at formulating the ethics of an endangered (and dangerous) life in *SE* by exploring different human types' engagement with danger. Nietzsche casts the Rousseauian type as depraved and disenfranchised. Resenting his inferiority, he rashly springs into violent action, succumbing to his peculiar danger of degenerating into a hedonistic savage (*SE* 4, 151). The Goethean type, in turn, is a distant observer, "a preservative and conciliatory power—but with the danger [. . .] that he may degenerate into a philistine" (*SE* 4, 152), i.e., an uncritical mouthpiece of shallow public opinion.

Nietzsche has little interest in those types, because they lack the distinctive trait that he values in the Schopenhauerian one: the ability to recognize and withstand the dangers that assail him (*SE* 3, 137). For Nietzsche, the Rousseauian and the Goethean types represent abortive attempts at raising themselves to a fully human level (*SE* 5, 159). In Katsafanas's terminology, Nietzsche poses the challenge of achieving a unified self whose instincts are in

harmony with its capacity for self-conscious choice (2014: 185-216).¹⁹ Neither giving in to its urges nor assuming that it fully controls its actions, an individual's unity derives from a stable attitude in light of revelations of deeper, unconscious motivations (Katsafanas 2016: 165). For Nietzsche, a Schopenhauerian education strives for “a harmonious whole [. . .], a simultaneous sounding of many voices in one nature [in which] everything, knowledge, desire, love, hate, strives towards a central point, a root force” (*SE* 2, 131). The Schopenhauerian type is the only one capable of achieving this unity, precisely because he is courageous enough to confront his dangers and heed Nietzsche's exhortation, “we want to be the true helmsman of [our] existence and refuse to allow our existence to resemble a mindless act of chance. One has to take a somewhat bold and dangerous line with this existence” (*SE* 1, 128).

Three “constitutional dangers” (*SE* 3, 144) threaten the Schopenhauerian type: his deeds risk being inauthentic and misunderstood (*SE* 3, 139–40), he must undergo extreme mental and physical trials in his quest for truth (*SE* 3, 140), and humanity constantly disappoints him (*SE* 3, 142). Thus, he might flee into solitude, despair, apathy, or madness. In addition, Nietzsche points to specific “dangers which arose from [Schopenhauer's own] *age*” (*SE* 3, 144): a desiccated historical sense with an ironic attitude of a closed future, and the cynical view that the past could not have been otherwise. While these threats are daunting, Nietzsche asserts that there has been at least one historical individual able to withstand them without degenerating—Arthur Schopenhauer himself. “It is, however, nothing less than a miracle that he was able to become this human example: for he was pressed upon, from within and without, by the most tremendous dangers which would have crushed or shattered any weaker being” (*SE* 3, 137). An ethics of danger is not impossible, but it requires courage and fortitude. It is not a coincidence or an oversight that Nietzsche does not lay out a positive

¹⁹ In a similar vein, cf. the distinction between the minimal self and the normative self in Anderson.

ethics of danger in the form of specific moral prescriptions, but only indicates an ideal type by instructing us what *not* to do: not to deceive oneself or others, to inquire into and become one's own true self and to take full responsibility for it (*SE* 4, 152).²⁰ Nietzsche's ethical interests in *SE* run much deeper than providing a positive ethical program: the Schopenhauerian exemplar puts upon *us* the onus of confronting our dangers and acquiring our own coherent self from which to draw the criteria for our actions. A pre-given set of instructions would rule out such a dangerous exposure.

To be sure, such an undertaking could fail even more colossally than the undertakings of the doomed Rousseauian and Goethean types. Only the Schopenhauerian type can even begin to take on the daunting ethical question in the more radical sense in which Nietzsche conceives it: to uncover and grapple courageously with the dangers that beset us, i.e., to face the paradoxical, self-contradictory core of everything that we hold to be stable and coherent.

Nietzsche's philosophy of danger advances a portentous ethical proposition: only those who have the courage, fortitude, and truthfulness to endanger themselves stand a chance of being ethical, while success in this endeavor is all but guaranteed. The outcome of such an exposure is potentially catastrophic: Nietzsche concludes *SE* with the remark that, upon the arrival of a Schopenhauerian exemplar, "all things are at risk [in Gefahr]. It is as when a conflagration has broken out in a city, and no man knows what is safe, or where it will end" (*SE* 8, 193). In his middle period works, Nietzsche will develop the full implications of the dangerous wager he offers. Old morality will be debunked as never even having reached the threshold of genuine ethics, genuine ethics itself might turn out to be impossible, and the ethically inquiring human being will constantly have to court and contend with danger.

²⁰ This motif foreshadows Nietzsche's later theme of the free spirit's intellectual honesty, or *Redlichkeit*.

3. Embarking on Perilous Seas: The Demise of Metaphysics and the Ethical Question

Franco calls *HH* (1878/1880), *D* (1880), and *GS* I-IV (1882) the “free spirit-trilogy,” in reference to the entity that takes center stage in Nietzsche’s philosophizing around that time (Franco). The free spirit moves along perilous paths. He is the restless, solitary, inquiring antipode of the complacent, socially immersed philistine who fanatically holds on to his dogmatic convictions even when they have been refuted (Reginster 2003). In his 1886 prefaces, Nietzsche characterizes his own philosophizing as a free spirit’s solitary, dangerous wandering. In *D*, he says, “that is inherent in ‘proceeding on one’s own path’: [...] all the perils [Gefahr], accident, malice and bad weather which assail him he has to tackle by himself” (*D* P:2). Similarly in *GS*: “we daredevils of the spirit who have climbed the highest and most dangerous peak of present thought and [...] have looked *down* from there” (*GS* P:4) (see also *HH* I P:4 and *HH* II P:2). Words such as “Gefahr,” “(un)gefährlich,” “Gefährlichkeit,” and “(un)gefährdet” occur frequently throughout the middle works, and Nietzsche’s aphoristic style takes greater risks with regard to argumentative rigor and clarity. Danger seeps into the very fabric of the free-spirit works of Nietzsche’s middle period. The free spirit brings danger, is in danger, and courageously seeks it out. He has a deep commitment to intellectual honesty, which drives him to tease out the inherent contradictions of metaphysical thinking. The ethical question fully presents itself in moments of metaphysical breakdown, glimpsed in situations of severest danger.

The free spirit brings danger to established dogmas by incessantly trying to dismantle metaphysics, i.e., the Platonic dualism between unconditioned, permanent, essential truths and their derivative, contingent counterparts (Thorgeirsdottir). The first chapter of *HH* I, entitled “Of First and Last Things,” for instance, attacks notions such as the thing in itself (*HH* I:16),

substance (*HH* I:19), free will (*HH* I:18), and causation (*HH* I:13). Similar critiques abound in *D* (e.g., *D* 10, 37) and *GS* (e.g., *GS* 110, 112, 127). Moreover, Nietzsche disputes not only specific metaphysical concepts, but also their underlying assumptions of constancy (*HH* I:21), the applicability of generals to particulars (*HH* I:19), the separateness of identities (*HH* I:1), and truth as correspondence (*HH* I:11). The agent that exposes these underlying fallacies and thereby endangers Platonic metaphysics is the free spirit, for whom *HH* is conceived: it is a “book for free spirits.”

Indeed, Nietzsche goes so far as to argue that metaphysics *itself* is an error and illusion. Metaphysics is not simply riddled with corrigible fallacies. The real fallacy consists in having conceived of a dualist metaphysics that is fundamentally self-contradictory. A paradigmatically condensed argument occurs in the unpublished, slightly earlier *TL*, which features a precursor to the free spirit under the name of “liberated intellect [freigewordener Intellekt]” (*TL* 152). Nietzsche argues that humans cannot come together into a cohesive, life-preserving community unless they standardize the usage of language (*TL* 143). Truth as correspondence emerged historically when the “correct” utilization of linguistic markers became arbitrarily fixed, obscuring the fact that, originally, utterances were highly particular metaphors, i.e., non-descriptive, linguistic expressions that were translations of physical stimuli into mental images and sounds (*TL* 143). “[T]he contrast between truth and lying comes into existence here for the first time” (*ibid.*). Over time, the particularity of those expressions was forgotten, as they became widely applied as abstract, general concepts within a systematic framework (*TL* 145–46). Metaphysics and its assumption of enduring essences are illusions that originated in a linguistic trick: to achieve a sense of security, a group feigned a stable, predictable world by fixing a certain use of language (*TL* 147). A crucial prerequisite for such a protective framework to function is that *everyone* uses language in the prescribed way, which is

accomplished by framing it as a *moral* imperative not to deceive anyone (*TL* 143). Metaphysics emerges as a morally sanctified set of dogmas on how to employ language “correctly”: to be good is to be truthful, and to be truthful is to be good.

The free spirit (or liberated intellect) sees through this trick of language. Unfettered from a rigid dogmatism, he plays with concepts and disregards standard rules: this “vast assembly of beams and boards to which needy man clings, thereby saving himself on his journey through life, is used by the liberated intellect as a mere climbing frame and plaything on which to perform its most reckless tricks” (*TL* 152). Free spirits endanger the established order by undermining both particular metaphysical concepts and Platonic metaphysics as such. As Berry emphasizes, their free play with concepts points to the “danger and rashness of our making pronouncements about the nature of a reality behind the appearances” (Berry 66), and their curiosity and courage renders them far less dependent on the security afforded by metaphysical frameworks than those fanatical philistine spirits who fear for their established ways of life and are bound by their metaphysical need (*TL* 152–53).

Free spirits not only endanger systematic speculative theories of ultimate truths, but also the rules of conduct for everyday practice. This comfortable life in the group is crucially protected by herd morality, which consists of rigid behavioral rules established at a time when life was much less certain (*D* 9; *HH* I:96). Those early moral systems strategically lauded conforming behavior and condemned deviation (*D* 9): violating conventional morality “is dangerous, and even more injurious to the *community* than to the individual” (*HH* I:96). For herd morality, there is essentially one “danger of dangers—the individual” (*D* 173).

Such moral systems—of which Christian morality is an outgrowth for Nietzsche—claim to rest on unchanging metaphysical truths. The free spirit, however, discerns that those systems reinforce merely the *illusion* of such truths. False group morality and erroneous

metaphysics mutually bolster each other. On one side, morality declares metaphysics' last truths to be sacrosanct: "the *mores* of morality sustained the faith that all of man's inner life was attached to iron necessity with eternal clamps" (*GS* 46). In return, metaphysics props up a historically contingent herd morality with the markers of permanence and essential truth. As a result, "a false ethics is erected, religion and mythological monsters are then in turn called upon to buttress it, and the shadow of these dismal spirit in the end falls across even physics and the entire perception of the world" (*HH* I:37). The true is the good, and the good is the true. When the free spirit attacks the foundations of metaphysics, he simultaneously attacks herd morality. He is at once dangerous and "evil."

The means by which the free spirit unravels herd morality is by rigorously following its own most central imperative: being truthful. As Nietzsche stresses, the free spirit's defining characteristic is his unswerving commitment to intellectual honesty (*Redlichkeit*), i.e., his commitment not to deceive others or himself (*Gau* 308-09). Unlike the philistine fanatic, he is not restrained by a cowardly demand for certainty that relies on articles of faith to block genuine inquiry (Reginster 2003, 60, 81), but maintains a deep curiosity and penchant for uncertainty (Reginster 2013, 458). Insisting on rigorous thinking, the free spirit teases out a thought's consequences, no matter how paradoxical or unpalatable they might be (*D* 101; *GS* 2). The distinctive Schopenhauerian trait has become the core attribute of Nietzsche's philosophizing entity in his middle period. Just as it was explored in *DS*, *HL*, and *SE*, the strategy consists not in rejecting the adversary's reasoning, but in embracing it to the fullest extent. In his commitment to intellectual honesty, the free spirit adopts herd morality's core principle not to deceive or be deceived more radically than any member of that group. Precisely this radical commitment takes the free spirit to the paradoxical insight that morality itself is based on an erroneous metaphysics, i.e., on a deception. If it is imperative for the

moral person not to deceive, but morality is grounded in deceptions, then anyone taking this moral imperative seriously has to conclude that a morality based on metaphysics cannot be truly moral. Intellectual honesty can then only call for metaphysics' abdication. The free spirit performs the tragically heroic insight: "faith in morality is withdrawn—but why? *Out of morality!*" (D P:4).

The free spirit does not just bring danger to dogmatic worldviews. He, too, is beset by dangers, as he is not exempt from the uncertainties that result from the crumbling of metaphysical and moral frameworks. Reminiscent of the dangers of the human types in *SE*, such dangers include the temptation to fall back into a caricature of a moralist (*GS* 107), or to become a hedonistic savage (*HH* I:542) or apathetic nihilist disavowing the possibility of truth altogether (*HH* I:29). Giving in to these temptations is dangerous, not just because they lead to degenerate behavior, but also because they result from degenerate (i.e., careless) reasoning: Nietzsche insists that metaphysically-grounded group moralities are erroneous, but this does not refute the possibility of morality altogether. Since faulty reasoning does not allow for *any* defensible conclusion (*HH* I:30), the free spirit has to hold on to his commitment to intellectual honesty, remain skeptical, suspend (Berry chs. 1, 6), and navigate his way between the Scylla of dogmatism and the Charybdis of nihilism. This perilous condition of having to chart his own path without any foregone conclusions causes Nietzsche to characterize the free spirit as the happiest and most vulnerable being: "For us life is more dangerous: we are made of glass; woe unto us if we merely *bump* ourselves! And all is lost if we *fall*!" (*GS* 154).

As Nietzsche's philosophy of danger reveals the mutual exclusivity of a Platonic metaphysics and a genuine ethics, it has raised its stakes to being an honest practice of philosophizing and living. This is where the ethics of the free spirit emerges in a radically new, non-dogmatic, and non-positive sense: it is an ongoing, committed practice of uncovering

metaphysical illusions by penetrating to the paradoxical, self-undermining assumptions that futilely seek to ground those illusions. Berry captures this essential characteristic of Nietzsche's philosophy as the Pyrrhonian's task to "keep the spirit of investigation alive...by not coming to rest with a judgment one regards as true," (Berry 36) whether this judgment be positive (dogmatic) judgment, or negative (nihilistic). Cashed out in ethical terms, this means that the free spirit's conduct remains ethical as long as he musters the courage to uphold his commitment to uncovering deceptions, even about himself. He therefore embodies a paradox: his ethics consists precisely in the sustained process of *trying to find* an ethics without ever coming to rest in a dogmatic theory, a grueling "mode of existence [that] is not for the faint of heart" (Berry 210).

The parable entitled, "Casuistical" (*D* 436) captures the free spirit's unsettling situation when his ethical commitment leads to a collapse of metaphysics and herd morality, which does *not* lead to moral nihilism. Rather, it causes a keen sense of ethical urgency, pointing to the new kind of ethics Nietzsche proposes. The passage deserves to be cited in full:

There is a wicked dilemma to which not everyone's courage and character are equal: as a passenger on a ship to discover that the captain and steersman are making dangerous mistakes and that one is superior in nautical knowledge—and then to ask oneself: how if you should incite a mutiny against them and have them both seized? Does your superiority not obligate you [verpflichtet] to do so? And would they not also be in the right if they locked you up for undermining discipline?—This is a metaphor for more elevated and agonising situations: in which the question in the end is always what guarantees our superiority, and justifies our faith in ourself, in such cases. Success? But one has already had to *do* the thing which bears within itself all the dangers—and not only dangers for us but for the ship as well. (*D* 436, translation altered)

In one and the same moment, comforting dogmatic truths disappear and the ethical urgency to do the right thing arises, without its being at all clear what the right thing would be. The free spirit recognizes that everyone's sense of security is illusory, as the stewards tasked with steering the ship lack the knowledge to ensure its safety. The ship is at the mercy of the

unpredictable, hostile sea. There is no solid ground of metaphysics, only the deceptive support of the deck of a ship in peril of going under. Given the stewards' incompetence, it is only a matter of time before the catastrophe occurs, which raises the pressing ethical question: for the sake of everyone's lives, what should the free spirit do?

One thing he cannot do in good conscience is to return to being a deluded passenger. The question, "does your superiority in knowledge not obligate you to take action?" gives him no rest. Responses such as trusting the crew, becoming apathetic or hedonistic, or going mad evade this question and the danger at hand. The clear-sighted, ethically concerned free spirit can no longer rely on established doctrines in deciding what to do. The guiding principles supposedly protecting society turn out *not* to guide and protect it, which releases him from being bound by them. He has to decide on the right thing to do by himself.

This release does *not* unburden him, but weighs him down with a dilemma: not intervening is evil because it abandons everyone to the danger they fail to see; yet, intervening is evil, as it might cause a turmoil that jeopardizes everyone's safety even more. The dilemma is "wicked," as the free spirit has to choose between two evils. His superior knowledge does not guarantee safety, or a clear plan for action. No matter how sophisticated his nautical skills, to dispel the threat posed by the sea entirely and indefinitely is impossible, for one never knows whether they will be sufficient in future storms. Whichever choice he makes risks hastening the ship's going under.

It is clear, however, that the only one on the ship even able to engage the ethical question is the free spirit. Everyone else is too deluded to recognize how their supposed morality is in violation of the very function it allegedly serves—preserving the life of the group. No matter how dire the implications, an intellectually honest free spirit (and we ourselves) can no longer resort to a metaphysically-grounded morality of the herd, as it forecloses the deep,

personal struggle that has to undergird any genuinely ethical behavior. *The genuine question of ethics appears in all its urgency precisely when free-spiriting truthfulness reveals metaphysics to be self-contradictory, and hence an illusion.* Out at sea, metaphysically unmoored, and in moral chaos, the free spirit needs more than anything the courage to prevail against a hostile crowd and a hostile sea, making a decision that depends entirely on him, and that may turn out to be calamitous.

4. Concluding Remarks: Danger's Strategic Role in Nietzsche's Philosophy

It is not a coincidence that Nietzsche frequently associates danger with the metaphor of a ship out at sea, not just in *D* 436, but throughout his middle works. The metaphor variously involves optimistic calls to set sail into the unknown (*HH* I:477; *D* 432, 575; *GS* 283, 343, 382), encouragements to persevere in a storm (*HH* I:9; *GS* 318), warnings not to trust the menacing calmness of the water (*HH* II:P:1; *HH* II:WS 243), and injunctions to hold back the creeping despair at being lost on an infinite ocean (*HH* I:248; *HH* II:WS 179; *GS* 124). This multiplicity of moods indicates the personal tribulations that a free-spirited philosopher of danger has to endure. Throughout the metaphor's middle period iterations, however, there is no mention of reaching solid ground again. While Nietzsche's ship still reaches new shores after its perilous discoveries in *HL* (10, 116), this possibility disappears in the more anti-metaphysical middle works. Once we embark, we are out at sea for good; when we are in *Gefahr*, we are in *Ge-fahr*, i.e., constant motion.

Indeed, the German word *Gefahr* derives from a root combining a cluster of meanings that cannot have been lost on Nietzsche the philologist. The Middle High German *gevāre*, means *Hinterlist* (a backhanded ploy) or *Betrug* (stratagem, deceit), dating back to the older *fāra*, which signifies *Nachstellung* (chasing, going after) (Kluge 305, "Gefahr"). *Fāra*, as well as the more basic root word *far-a* (meaning "fahren" as being in any kind of motion (Kluge 245,

“fahren”), both derive from the Indogermanic root **per-*, meaning “going across” (ibid.). *Gefahr* is thus rooted in a semantic cluster that combines unpredictability, deception, threat, pursuit, motion, and the inability to come to rest. Danger does not afford us the comfort of staying still, for there is always something coming after us. Staying in motion is the only option.

This semantic cluster reinforces the structural dynamic of the notion of danger that Nietzsche employs, indicating how a practice of free-spirited philosophizing could be activated among his readers. After all, if his philosophy of danger leads to such unsettling insights as he discovers, why follow Nietzsche’s invitation to embark on it at all? Why awake from the slumber of a ready-made metaphysics and herd morality if, after an initial stage of joyful subversion, only the gaping abyss of nihilism seems to remain? Nietzsche’s answer is that we are *already* on the ship. The danger of going under already threatens all of us, whether we see it or not. Intellectually honest free-spirit philosophizing is not one of several options, but an inevitable, grueling ethical imperative. Danger functions as a strategic operator in Nietzsche’s thought that aims to *activate* this philosophizing and the non-dogmatic ethical stance towards truth that comes with it.

Everything in Nietzsche’s philosophy of danger therefore depends on whether he can drive home danger’s urgency and elusiveness. Danger must not be misunderstood in the terms of the Platonist metaphysical dualism of essence and appearance. On the one hand, treating danger as a mere appearance without a true reality behind it would give false license to ignoring it. On the other hand, treating danger as an essential truth about the world would raise the equally false impression that, once it has been identified we have mastered it and it ceases to threaten us. Instead of falling on either side of the Platonic metaphysical dualism, danger unceasingly calls the dualism into question. Precisely this undecidable status makes danger so self-effacing, uncontrollable, ubiquitous, and potent, leaving us with only one choice: to

engage it vigilantly and undauntedly, wherever it may lurk. Danger gives us no rest, as it continually draws us out to sea, where we have to stay in motion lest we go under in a betrayal to our commitments to intellectual honesty and a genuine ethical practice. We are *Gefahr* from the very beginning, and Nietzsche's philosophy of danger puts us in motion, into *Ge-fahr*.

Such an approach to philosophy becomes fully intelligible in the interstitial space between the Anglophone and postmodern modes of engaging Nietzsche. Reconstructing systematic frameworks of Nietzschean ethics or metaethics under the headings of perfectionism, virtue ethics, evolutionism, or any other moral framework in the service of a flourishing life may yield plausible results, but misses the deeper point that Nietzsche sought to grapple with ethics precisely from the angle of any such secure systematicity *breaking down*. A Nietzschean ethics should not be sought in efforts to counter, alleviate, resolve, or ignore the paradoxes in his thinking, but has to take them as its starting point. Yet this approach must not lead into the postmodernists' trap of sidestepping the question of a Nietzschean ethics altogether. In fact, the breakdown of metaphysics makes the question of a genuine ethics urgent and unavoidable. Nietzsche's central occupation with ethics and his paradoxical thinking must not be separated, despite its dangerous (and potentially tragic) implications.

WORKS CITED

Nietzsche's Works

- BT *The Birth of Tragedy*, in *The Birth of Tragedy and Other Writings*, ed. Raymond Geuss and Ronald Speirs, trans. Ronald Speirs. Cambridge: Cambridge University Press, 1999. 1-116. Print.
- BGE *Beyond Good and Evil: Prelude to a philosophy of the future*, ed. R. P. Horstmann and J. Norman, trans. J. Norman. Cambridge: Cambridge University Press, 2002. Print.
- CW “The Case of Wagner: A Musician’s Problem.” In: Friedrich Nietzsche. *The Anti-Christ, Ecce Homo, Twilight of the Idols, and Other Writings*, ed. A. Ridley and J. Norman, trans. J. Norman. Cambridge: Cambridge University Press, 2005 [1888]. 231-262. Print.
- D *Daybreak. Thoughts on the Prejudices of Morality*, ed. Maudemarie Clark and Brian Leiter, trans. R. J. Hollingdale. Cambridge: Cambridge University Press, 1997. Print.
- DS “David Strauss, the Confessor and Writer.” In: Friedrich Nietzsche. *Untimely Meditations*. ed. Daniel Breazeale, trans. R. J. Hollingdale. Cambridge: Cambridge University Press, 1997 [1873]. 1-56. Print.
- EH *Ecce Homo: How to Become What you Are*. In: Friedrich Nietzsche *The Anti-Christ, Ecce Homo, Twilight of the Idols, and Other Writings*. ed. A. Ridley and J. Norman, trans. J. Norman. Cambridge: Cambridge University Press, 2005. 69-153. Print.
- GM *On the Genealogy of Morality*. ed. Keith Ansell-Pearson, trans. Carol Diethe. Revised student edition. Cambridge: Cambridge University Press, 2008 [1887]. Print.
- GS *The Gay Science. With a Prelude in Rhymes and an Appendix of Songs*, trans. Walter Kaufmann. New York: Vintage Books, 1974. Print.
- HL “On the Uses and Disadvantage of History.” In: Friedrich Nietzsche, *Untimely Meditations*, ed. Daniel Breazeale, trans. R. J. Hollingdale. Cambridge: Cambridge University Press, 1997 [1874]. 57-124. Print.
- HH *Human, All-Too-Human*, trans. R. J. Hollingdale. Cambridge: Cambridge University Press, 1996. Print.
- SE *Schopenhauer as Educator*, In: *UM*, 125-194. Print.
- TL “On Truth and Lying in a Non-Moral Sense,” In: *BT*, 139-153. Print.

UDH *On the Uses and Disadvantages of History for Life*, In: Friedrich Nietzsche. *Untimely Meditations*. ed. Daniel Breazeale, trans. R. J. Hollingdale. Cambridge: Cambridge University Press, 1997 [1873]. 57-124. Print.

Other References

Anderson, R. Lanier. "What is a Nietzschean Self?" In: Janaway and Robertson, eds. *Nietzsche, Naturalism, and Normativity*, Oxford: Oxford University Press, 2011. 202-235. Print.

Behler, Ernst. "Nietzsches Studium der griechischen Rhetorik nach der KGW." *Nietzsche-Studien* 27 (1998). 1-12. Print.

Berry, Jessica N. *Nietzsche and the Ancient Skeptical Tradition*. Oxford: Oxford University Press, 2011. Print.

Breazeale, Daniel. "The Hegel-Nietzsche Problem." *Nietzsche-Studien* 4 (1975) 146-164. Print.

Brobjer, Thomas H. "Nietzsche's Knowledge of Marx and Marxism." *Nietzsche-Studien* 31 (2002). 298-313. Print.

Cauchi, Francesca. "Hegel and Nietzsche on Thought, Freedom and 'the Labour of the Negative'." *Journal of European Studies* 46.2 (2016). 1-16. Print.

Conway, Daniel W. *Nietzsche's Dangerous Game: Philosophy in the Twilight of the Idols*, Cambridge: Cambridge University Press, 1997. Print.

de Almeida, Rogerio Miranda. *Nietzsche and Paradox*, trans. Mark S. Roberts. Albany: State University of New York Press, 2006. Print.

de Man, Paul. "Rhetoric of Tropes (Nietzsche)." In: Paul de Man. *Allegories of Reading. Figural Language in Rousseau, Nietzsche, Rilke, and Proust*. New Haven and London: Yale University Press, 1979. 103-118. Print

de Man, Paul. "Rhetoric of Persuasion (Nietzsche)." In: Paul de Man, *Allegories of Reading. Figural Language in Rousseau, Nietzsche, Rilke, and Proust*. New Haven and London: Yale University Press, 1979. 119-131. Print.

Deleuze, Gilles. *Nietzsche and Philosophy*, trans. H Tomlinson. London: The Athlone Press, 1983 [1962]. Print.

Dellinger, Jakob. "Vernichtung, Grausamkeit, Gefahr." In: *Nietzsches Philosophie des Unbewussten*, ed. Jutta Georg, Claus Zittel. Berlin: de Gruyter, 2012. 233-242. Print.

- Derrida, Jacques *Spurs. Nietzsche's Styles / Éperons. Les Styles de Nietzsche*, trans. Barbara Harlow. Chicago and London: University of Chicago Press, 1979 [1978]. Print.
- Dudley, Will. *Hegel, Nietzsche, and Philosophy. Thinking Freedom*. Cambridge: Cambridge University Press, 2002. Print.
- Forster, Michael. "Hegel's Dialectical Method." In: F. C. Beiser, ed., *The Cambridge Companion to Hegel*. Cambridge: Cambridge University Press, 1993. 130-170. Print.
- Foucault, Michel. "Nietzsche, Genealogy, History." In: Michel Foucault. *Language, Counter-Memory, Practice. Selected Essays and Interviews*, ed. Donald F. Bouchard, trans. Donald F. Bouchard and Sherry Simon. Ithaca: Cornell University Press, 1977. 139-164. Print.
- – –. "Nietzsche, Freud, Marx." In: Michel Foucault. *Aesthetics, Method, and Epistemology*. ed. James D. Faubion, trans. Robert Hurley et al. New York: The New Press, 1998. 269-278. Print.
- Franco, Paul. *Nietzsche's Enlightenment: The Free-Spirit Trilogy of the Middle Period*. Chicago: University of Chicago Press, 2011. Print.
- Frischmann, Bärbel. "'Gefährlich leben!'" In: *Nietzsche-Lexikon*, ed. Christian Niemeyer. Darmstadt: Wissenschaftliche Buchgesellschaft, 2009. Print.
- Grau, Gerd-Günther. "Redlichkeit, intellektuelle," In: *Nietzsche-Handbuch: Leben – Werk – Wirkung*, ed. Hans Ottman. Stuttgart: J.B. Metzler, 2000. Print.
- Holz, Hans Heinz. *Dialektik: Problemgeschichte von der Antike bis zur Gegenwart. Band V: Einbeit und Widerspruch III. Die Ausbreitung der Dialektik*. Darmstadt: Wissenschaftliche Buchgesellschaft, 2011. Print.
- Houlgate, Stephen. *Hegel, Nietzsche and the Criticism of Metaphysics*. Cambridge: Cambridge University Press, 1986. Print.
- Jaspers, Karl. *Nietzsche: An Introduction to the Understanding of His Philosophical Activity*, trans. C. F. Wallraff and F. J. Schmitz. Tucson: The University of Arizona Press, 1965 [1936]. Print.
- Joisten, Karen. "Das Gefährliche der Gefahr," In: Karen Joisten, *Die Überwindung der Anthropozentrität durch Friedrich Nietzsche*, Würzburg: Königshausen & Neumann, 1994. 70-97. Print.
- Kaufmann, Walter. *Nietzsche: Philosopher, Psychologist, Antichrist*. 4th ed. Princeton: Princeton University Press, 1974 [1950]. Print
- Katsafanas, Paul. *The Nietzschean Self: Moral Psychology, Agency, and the Unconscious*. Oxford: Oxford University Press, 2016. Print.

- – –. “Nietzsche and Kant on the Will: Two Models of Reflective Agency,” *Philosophy and Phenomenological Research* 89.1 (2014). 185-216. Print.
- Kluge Friedrich. “Gefahr.” In: *Etymologisches Wörterbuch der deutschen Sprache*. 23. erweiterte Auflage, bearb. Elmar Seebold. Berlin: de Gruyter, 1995. Print.
- Löwith, Karl. *From Hegel to Nietzsche. The Revolution in Nineteenth-Century Thought*, trans. David E. Green. New York and London: Garland Publishing, 1984 [1941]. Print.
- Marx, Karl. *Capital. A Critique of Political Economy*. ed. Friedrich Engels. New York: The Modern Library, 1906 [1872]. Print.
- – –. “A Contribution to the Critique of Hegel’s Philosophy of Right. Introduction.” In: Karl Marx. *Early Writings*. trans. Rodney Livingstone and Gregor Benton. New York: Vintage Books, 1975 [1843]. 243-258. Print.
- Meakins, William. “Nietzsche, Carlyle, and Perfectionism,” *The Journal of Nietzsche Studies* 45.3 (2014). 258-278. Print.
- Müller-Lauter, Walter. *Nietzsche: His Philosophy of Contradictions and the Contradictions of His Philosophy*, trans. David J. Parent. Berlin and New York: Walter de Gruyter, 1999 [1971]. Print.
- Moore, Gregory. *Nietzsche, Biology and Metaphor*. Cambridge: Cambridge University Press, 2002. Print.
- Nehamas, Alexander. *Nietzsche: Life as Literature*. Cambridge and London: Harvard University Press, 1985.
- Pasley, Malcolm. “Nietzsche’s Use of Medical Terms.” In: Malcolm Pasley, ed., *Nietzsche: Imagery and Thought*. London: Methuen, 1978. 123-158. Print.
- Reginster, Bernard. “What Is a Free Spirit? Nietzsche on Fanaticism,” *Archiv für Geschichte der Philosophie* 85.1 (2003) 51-85. Print.
- – –. “Honesty and Curiosity in Nietzsche’s Free Spirits,” *Journal of the History of Philosophy* 51.3 (2013) 441-463. Print.
- Renzi, Luca. “Das Ohr-Motiv als Metapher des Stils und der ‘Zugänglichkeit.’” *Nietzsche-Studien* 26 (1997) 331-349. Print.
- Robling, Franz-Hubert. “Plastische Kraft. Versuch über rhetorische Vernunft bei Nietzsche.” *Nietzsche-Studien* 25 (1996) 87-98. Print.
- Ruehl, Martin A. “Das Allgemeine und sein Bild: Zur Geschichtsphilosophie Jacob Burckhardts.” *Historische Zeitschrift* 296.1 (2013) 49-83. Print.

- Schacht, Richard. "Hegel, Marx, Nietzsche, and the Future of Self-Alienation." *The International Journal of Sociology and Social Policy* 11.6 (1991) 1-16. Print.
- Simonis, Linda. "Der Stil als Verführer. Nietzsche und die Sprache des Performativen." *Nietzsche-Studien* 31 (2002) 57-74. Print.
- Stegmaier, Werner. "Nietzsches Herausforderung an seine Leser: Gefährlich befreiendes Philosophieren." In: *Nietzsches Befreiung der Philosophie. Kontextuelle Interpretation des V. Buchs der Fröhlichen Wissenschaft*. Berlin: de Gruyter, 2012. 3-24. Print.
- Thorgeirsdottir, Sigridur "Metaphysik," In: *Nietzsche-Handbuch: Leben – Werke – Wirkung*, ed. Hans Ottman. Stuttgart: J.B. Metzler, 2000. 281-283. Print.
- Widmann, Joseph Victor. "Nietzsches gefährliches Buch," *Der Bund* 37.256-257 (Bern: September 16 and 17, 1886); cited in Hauke Reich, *Rezensionen und Reaktionen zu Nietzsches Werken 1872–1889*, ed. Günter Abel and Werner Stegmaier. Berlin: de Gruyter, 2013. 618-623. Print.
- Williams, Robert R. *Tragedy, Recognition, and the Death of God. Studies in Hegel and Nietzsche*. Oxford: Oxford University Press, 2012. Print.
- Wood, Allen W. "Hegel and Marxism." In: F. C. Beiser, ed., *The Cambridge Companion to Hegel*. Cambridge: Cambridge University Press, 1993. 414-440. Print.
- Zimmermann, Robert L. "On Nietzsche." *Philosophy and Phenomenological Research* 29.2 (1968). 274-281. Print.
- Zitko, Hand. *Nietzsches Philosophie als Logik der Ambivalenz*. Dissertation, Frankfurt. Published Würzburg, 1992 [1989]. Print.
- Zittel, Claus. *Selbstaufhebungsfiguren bei Nietzsche*. Würzburg: Königshausen und Neumann, 1995. Print.

CHAPTER 2: THE INDUCTIVITY OF THE LIFEWORLD IN HUSSERL'S LATE

THOUGHT

INTRODUCTION

In the opening pages of his 1936 *Krisis der europäischen Wissenschaften und die transzendente Phänomenologie*, Husserl declares: “[D]ie Krisis der Philosophie [bedeutet] die Krisis aller neuzeitlichen Wissenschaften ... [und eine] immer mehr zutage tretende Krisis des europäischen Menschentums selbst, in der gesamten Sinnhaftigkeit seines kulturellen Lebens” (10).²¹ With this remark, Husserl joined a ubiquitous debate. “Crisis” was the watchword of the time, extending from mathematics (Gödel’s Incompleteness Theorem had shattered Hilbert’s program to axiomatize mathematics four years earlier) to the natural sciences (Heisenberg’s uncertainty principle of 1927 demonstrated the impossibility of simultaneously determining a particle’s exact location and momentum) to philosophy, which had fallen into embittered disputes between positivism, psychologism, Neo-Kantianism, and Lebensphilosophie. Each of those schools reacted to the rupture between everyday life and the technical sciences in their own way. Positivism focused on pragmatic, logical deduction of true statements from accepted axioms and bracketed metaphysical questions of life and transcendental subjectivity; psychologism countered with attempts to provide a natural-scientific, materialist explanation of mental life; Neo-Kantianism opposed materialism by reviving the Kantian project of grounding subjectivity in an ideal, a priori structure of reason; Lebensphilosophie was deeply skeptical of rationalism and insisted that life was beyond its grasp.

²¹ All page numbers are from the Husserliana pagination of the 2012 Meiner edition. All emphases are in the original.

Edmund Husserl (1859-1938), a mathematician-turned-philosopher who sought to reconcile the sciences and the humanities throughout his career, intervened in those many-voiced debates with the ambitious aim of correcting all of the above schools, grounding the sciences in life and restoring their meaning, and uncovering the “*die historische Bewegung der Offenbarung der universalen, dem Menschentum als solchem ,eingeborene[n]’ Vernunft*” (13-14) by leading us into phenomenology, a mode of thinking that is always conscious of its grounding in life. Husserl develops his crucial concept of the lifeworld in the *Krisis* to pave the way toward this aim: the lifeworld grounds a living, historical, intersubjective practice from which science emerges as one of many cultural achievements.

There literature on Husserl’s concepts of life and the lifeworld is vast. For example, in his study of the *Krisis*, Dermot Moran describes the lifeworld “as a rich, multi-faceted notion with some apparently contradictory and paradoxical features that have puzzled even sympathetic commentators” (Moran 178). This confusion originates from the lifeworld’s dual status as both ontological and transcendental. On the one hand, Husserl treats it as an ontological ground (Boden) of living practice that provides the experiential and material conditions of possibility for the theoretical sciences; the lifeworld grounds them in an experiential, unreflected practice. On the other hand, the lifeworld functions as a transcendental guiding thread (Leitfaden) or horizon that leads beyond this primitive practice and into the sciences, and a historical critique of them.²²

The interwoven ontological and transcendental aspects of the lifeworld leave unclear what its status of being is, whether or not the sciences belong to the lifeworld, whether or not we can ever leave it. Hans Blumenberg seems to vacillate on the question. In “Theorie der

²² See also Ströker’s introduction to the Meiner edition (2012) of the *Krisis*, pp. XIX-XX.

Lebenswelt” and “Lebenswelt und Technisierung unter Aspekten der Phänomenologie,” he insinuates we do, while in “Lebenswelt und Wirklichkeitsbegriff,” he re-reads Plato’s allegory of the cave to suggest that our lifeworldly practice is a process of leaving the lifeworld that never ends—and thus never succeeds. Ernst Wolfgang Orth sees the lifeworld as a concept that necessarily leads into confusion. David Carr contends that “Husserl’s term ‘life-world’ involves [...] several different concepts at once” (372). Nam-In Lee agrees and argues that Husserl’s lifeworld is an inherently pluralistic concept that allows for several fields of phenomenology of the lifeworld. Ulrich Claesges views the lifeworld as an ambiguous concept that is beset by an irreducible tension as an ontological ground and a transcendental horizon:

Der Begriff der Lebenswelt bei Husserl ... ist ... ein *ontologisch-transzendentaler Zwitterbegriff*. Die auch durch wiederholte Darstellungen bisher nicht beseitigte Dunkelheit der Husserlschen Ansätze zu einer Theorie der Lebenswelt haben wohl letztlich darin ihren Grund. (Claesges 97)

While Claesges sees no avenue to resolve the rift, he emphasizes the importance of future research to do so: “[Es] müßte eine Methode entwickelt werden, die Korrelation von Lebenswelt und lebensweltlichem Erfahren ... zu thematisieren” (Claesges 101).

The answer, I argue, lays in plain sight in Husserl’s account. What has, to my knowledge, been largely overlooked, is the mode of acting and thinking that Husserl explicitly, emphatically, and repeatedly inscribes into the practice of life: his rich notion of induction. “Induktion” does not even appear in the index of the Meiner Verlag version until its 2012 edition. Early on, Eugen Fink called for an exposition of the operative terms in Husserl’s phenomenology (Fink). This call, however, never received much traction. Explicit studies on induction in Husserl are thus few and far between. Two exceptions are Levin and Palermo, who home in on the relationship between eidetic variation and induction. Levin maintains that “induction and eidetic variation are [best] conceived as continuous, and even complementary” (Levin 3), while Palermo insists on their fundamental difference (Palermo 70). Their

discussions run aground, however, because neither author tries to clarify what Husserl means when he speaks of induction. Instead, they both assume that induction is an enumerative procedure of empirically examining case after case, thus always being confined to finitude, incompleteness, and non-essential insight. Ram Adhar Mall began to grasp the deep significance of induction in Husserl's thought by foregrounding induction as "die Urtendenz der Erfahrung[, die] immer vollzogen [wird] in dem Sinne, daß die ganze Lebenspraxis sie voraussetzt, zweitens wird sie auf dem Grund der Erfahrung vollzogen" (Mall 34). Other authors mention Husserl's treatment of induction in passing (Merleau-Ponty 64-77, Derrida 117-21, Sepp 260-79).

Recent scholarship has sometimes touched on induction in Husserl's late thought, but without fully developing its implications. For example, Dagfinn Føllesdal and Michael Friedman reconstruct Husserl's intellectual development from the *Ideen* to the *Krisis* in order to assess the significance of the lifeworld in Husserl's philosophy. Føllesdal shows that the phenomenological concept of the horizon enshrines a thoroughly inductive experiential practice (Føllesdal 34). Friedman, too, mentions that "[w]ithin the life-world itself, we have inductive regularities and causal relations that guide us in this process of continual correction [of our knowledge]" (Friedman 105), and that a science *of* the lifeworld needs to consciously remain in "that intuitive mode in which it operates" (105-06). However, neither author works out the importance of this inductive undercurrent of the lifeworld.

This general failure to focus on induction in Husserl's thought is surprising, given that induction is a philosophical problem, a scientific practice, and a mathematical method of proof—all of which touch on areas of thought that Husserl was familiar with and sought to reconcile. This warrants a return to Fink's call for developing the operative terms in Husserl's thought by looking at the *Krisis* and the key role induction plays as an operative concept.

This chapter will attempt to do five things. 1) It will lay out Husserl’s rich account of life as being irreducibly inductive. It will then reconstruct how primitive induction turned into formalized mathematical induction, which, upon closer scrutiny, turns out to be a *deductive* mode of thinking. Retracing this shift will elucidate Husserlian induction’s inherent proclivity toward self-formalization. 2) This chapter will then show how the lifeworld has to guarantee the continued inductive, self-clarifying practice of life, and how the lifeworld has to always remain partially obscure, since making its structure fully transparent would ground life in a deductive frame and divest it of its inductivity. 3) Circling in on the motivational problem in Husserl’s account—how precisely ought we to think the transition from the natural attitude to the theoretical attitude?—this chapter will lay out the deep affinity between the thought of Husserl and David Hume. Indeed, the lifeworld can be conceived as a concept designed to answer Hume’s problem of induction. 4) This chapter will then locate the solution to the motivational problem in an unlikely source: Aristotle’s concepts of nous, epagoge, and exemplarity. 5) The chapter will conclude with a discussion of how the concept of exemplar differs fundamentally from the concept of example, and how Husserl’s lifeworld and rich notion of induction cannot be understood without a proper conception of the exemplar.

I. MATHEMATIZATION AND INDUCTION’S TENDENCY TO TRANSFORM INTO DEDUCTION

1. Umstellung der Einstellung: Three Moments

In his pivotal section 9 on Galileo and his significance for European thinking, Husserl frames his historical inquiry as a reconstruction of a shift from a naïve, natural attitude to a scientific, theoretical attitude (a “Umstellung der Einstellung”). He seeks to retrace how it could come about that, around the time of the emergence of modern science in the

Renaissance, it became obvious to assume that nature was a rational, entirely law-governed structure that could be predicted with the deductive calculative tools of mathematics. Husserl's aspiration is to retrace the motivation that led to this change, as he unambiguously states in the opening lines: "*Was ist der Sinn dieser Mathematisierung der Natur*, wie rekonstruieren wir den Gedankengang, der sie motivierte?" (20). Galileo is the exemplaric figure in an exemplaric study (section 9 as a whole) that seeks to understand the motivation underlying such a fundamental attitudinal shift.

Husserl's genetic account of modern science traces an organic series of three crucial moments: 1) the measurement of extended bodies becomes more precise through iterative improvement of the measurement technique; 2) from that iterative improvement there arises an infinite task to strive for perfect precision; this task dimly anticipates universal, ideal objectivities that this perfected technique seeks to capture; 3) eventually, the world is no longer understood as an assemblage of extended bodies, but of ideal objectivities.

The first moment Husserl identifies is that the sciences, as practiced by the ancient Greeks, applied mathematics to the immediately encountered surrounding world by measuring extended bodies. This art of measuring (*Messkunst*) was done without the assumption that nature was itself governed by mathematically precise laws. Quantifying the dimensions of extended bodies was merely seen as an approximation of nature. Holding a standardized instrument against the body to be measured always left a discernible margin of imprecision. Seeking precision by holding standardized measurement units against physical always exposed a remainder of imprecision, which had to be eliminated by developing more precise tools. This technique was inherently (albeit crudely) inductive. Through repeated improvements of the measurement tools, the measurement method became more and more precise and technicized. Applying the technique improved it. The first knowledge-seeking engagement of humanity

with the world was thoroughly inductive in two ways: iteratively approximating external bodies and *itself*.

Iterative measurement is an expression of an inductive attitude to the surrounding world. According to Husserl, the repetitive occurrence of certain phenomena gives rise to the common assumption that the world is governed by stable laws that are causal and can be inductively derived: “Dieser universale Kausalstil der anschaulichen Umwelt macht in ihr Hypothesen, macht Induktionen, macht Voraussichten hinsichtlich der Unbekanntheiten der Gegenwart, der Vergangenheit, der Zukunft möglich” (29). In a supplement to the Galileo section, Husserl is even more emphatic, emphasizing

daß in der Welt als Welt der strömenden Erfahrung ... eine allumspannende Induktivität, insbesondere als Kausalstil herrscht, daß man von jedem Seienden aus und seinen zur Gewissheit gekommenen Eigenschaften aus induktiv erwarten, danach induktive Schlüsse machen kann, Induktionen, die sich im allgemeinen bewähren — das ist der Erkenntnisstil in betreff der der erkenntnismäßigen Welt, der Welt aus Erfahrung...” (Hua VI, 350)

In other words, Husserl’s first major contention in the Galileo paragraph is that, in Greek antiquity, a thoroughly inductive mode of engagement with the world—the art of self-improving measurement—and a thoroughly inductive worldview made for a functional symbiosis in which there was neither need nor room for deduction.

This repeated, iterative measurement soon became an infinite task. The second moment that Husserl identifies is that, once this iterative improvement of the method has happened often enough, each time falling short of the precision for which it aims, it eventually becomes clear that this improvement can be repeated infinitely many times. This infinite task aims at an infinitely precise measuring technique and infinitely precise body to be measured. Due to its own inherent developmental thrust, Husserl argues, this inductive process undergoes a crucial transformation:

[Die Meßmethode] ist eine Kunst, nicht als fertige Methode, etwas fertig zu machen, sondern *zugleich Methode, ihre Methode immer wieder zu verbessern ...*; gewinnt aber das ‘Immer wieder’ den mathematischen Sinn des ‘in infinitum’” (*Krisis* 43; see also Hua VI, 40)

As the inductive practice of improving the measuring technique becomes infinitized, the technique begins to anticipate and aim at ideal objectivities:

Ohne von hier aus tiefer in die Wesenszusammenhänge einzudringen ... werden wir schon verstehen, daß sich von der Vervollkommnungspraxis her, im freien Eindringen in die Horizonte *erdenklicher* Vervollkommnung im ‘Immer wieder’, überall *Limes-Gestalten* vorzeichnen, auf die hin als invariante und nie zu erreichende Pole die jeweilige Vervollkommnungsreihe hinausläuft. (*Krisis* 23)

Those intended (or aimed-for) objects of perfect precision are ideal objects: they can be clearly conceived, but never as such be encountered in nature. “Die Meßkunst wird also zur Wegbereiterin der schließlich universellen Geometrie und ihrer ‘Welt’ reiner Limesgestalten” (25). Along with this infinitization of its task comes the universalization of its conceived applicability. This method no longer simply anticipates or aspires to measure particular bodies with infinite exactitude. Rather, its ideal becomes infinite exactness *as such*. The inductive practice and worldview begin to anticipate a unified, overarching ideal world, i.e. a world of ideal objects that is governed by universal, perfectly exact, mathematical laws. In Husserl’s words:

Vermöge der reinen Mathematik und praktischen Meßkunst kann man für alles dergleichen Extensionale an der Körperwelt eine *völlig neuartige induktive Voraussicht schaffen*, nämlich man kann von jeweils gegebenen und gemessenen Gestaltvorkommnissen aus unbekannte und direkter Messung nie zugängliche in zwingender Notwendigkeit “*berechnen*.” (31)

This leads to the third key moment in Husserl’s genealogy of inductive science, a point that is inextricably linked to the second. As this infinitization and universalization of induction occur, and as induction anticipates an ideal realm of exact and perfectly calculable entities, a jump happens: The anticipated, ideal world is taken as the actual one. Nature is no longer

merely *approximated* by mathematical laws, but the laws provided by mathematics are taken to be the laws of nature. Nature is itself mathematical. The general hypothesis of universal inductivity (41) loses its status as a hypothesis and is taken as attained, confirmed, and corroborated:

Soweit wir bisher gekommen sind, ist zunächst nur ein allgemeiner Gedanke gewonnen, präzise ausgedrückt, eine allgemeine *Hypothese*: daß eine *typische Induktivität* in der anschaulichen Welt herrsche, eine sich in jenen alltäglichen Erfahrungen ankündigende, aber eine in ihrer Unendlichkeit verborgene. Freilich, für Galilei war sie nicht als *Hypothese* verstanden. ... Es galt nun [mittels mathematische exakter Methoden] die eigenartige Induktivität der Erfahrungswelt systematisch zu erfassen, die in der Hypothese vorausgesetzt war. (*Krisis* 41; see also Hua VI, 37-38)

As the art of measurement becomes more precise, more successful, and seems to come ever closer to its anticipated world of ideal objectivities, an unnoticed but fundamental shift occurs: science forgets its thoroughly inductive nature and origins. The mathematical sciences are thus disconnected from life, as they are conceived under the aegis of *mathesis universalis*, which constantly aims to expand its “Bereich der ihr möglichen *deduktiven Schlußfolgerungen auf neue Tatsachen* der quantifizierten Natur” (*Krisis* 47). Deduction is able to constantly expand because, in the transition from primitive to artful induction, the lifeworld became draped by an “*Ideenkleid*” (51), or “*Kleid der Symbole*” (52), i.e. an attitude of seeing the world as a mathematical cosmos. This led to a “Verselbständigung der erweiterten formalen Logik” (46), a thinking that *conceives of itself* as not extraneously grounded: “Zum Wesen aller Methode gehört die Tendenz, sich in ihrer Technisierung zu veräußerlichen” (48). This thinking elides and forgets that all deductions are still subject to experimental verification, and in fact never become independent from their inductive grounding (50). Deductive *thinking* performs the Galilean substruction (“*Unterschiebung*”) of a conception of the world as a mathematical, deductive *structure*, blinding us to its underlying inductive practice: “Das Ideenkleid macht es, daß wir für *wahres Sein* nehmen, was eine *Methode* ist” (52).

Galileo's discovery of the new method of thinking thus coincides with an occlusion: "Das ist alles Entdeckung-Verdeckung, und wir nehmen das bis heute als schlichte Wahrheit" (53). It is at this elusive juncture that the switch from natural to theoretical attitude happens. The scientists' interest shifts from inductive inquiry to the technocratic handling and deriving of formulas. The world is treated as a closed axiomatic system under the aegis of a *mathesis universalis*. This paradoxical moment of both self-realization and self-loss is what Husserl characterizes as the "entdeckende-verdeckende Moment." Husserl's historical account in sections 9 (a) through (g) of inductive science following its own formalizing tendencies thus traces how *induction—as a method and an attitude—is in and of itself drawn to deduction*.

2. Vignette: The Genesis of Mathematical Induction as an Exemplar for the Attitudinal Shift

How could induction turn into deduction? How could a non-scientific life practice, in touch with its surrounding world, precipitate itself into a technicized method that obscures its own origins? To gain a closer understanding of how such an attitudinal shift can occur, I suggest turning our attention to the purest, most exemplary artful induction there is: mathematical induction.

While philosophical accounts of induction have proliferated since Hume, the mathematical version is clearly defined (see Vickers 1.1). Suppose you want to prove proposition P of a certain collection C of infinitely many objects. First, assume that those objects are numbered 1, 2, 3, and so on, to infinity. The key step consists in proving the "inductive hypothesis": *if* some proposition P holds for *some* member (numbered n) of the collection, then P *also* holds for the following member (numbered $n + 1$) in that collection. Proving the inductive hypothesis thus does not prove P for any particular member, but establishes a relation between them: *If* it is the case that P is true for one member, it is also

true for the next member in the sequence. Next, one then establishes the “base case”: P does in fact hold for the first member $n = 1$. By the inductive hypothesis, P then also holds for $n = 2$. But by the Inductive Hypothesis, P then also holds for $n = 3$, and so on for all members of C . Once a single case has been demonstrated, the inductive hypothesis instantaneously kicks in infinitely many times, feeding its logical results back into its machinery. Mathematical induction makes infinity manageable by making true statements about an infinity of objects, and it is logically irrefutable.

A closer look reveals, however, that the name “mathematical induction” belies its logical architecture. The inductive hypothesis is an “if—then” statement. There is no actual stepwise process of progressive verification, generalization, self-correction, or idealization: the Inductive Hypothesis demonstrates P by way of a syllogistic inference that. *Mathematical induction is in fact deductive* (see Vickers, Section 1.1). Of course, mathematical induction and its deductive logical architecture did not spring into existence *ex nihilo*. Rather, they emerged historically from a genuinely inductive philosophico-mathematical mode of reasoning that morphed into its modern mathematical variant. This transformation parallels, and will clarify, the Galilean shift of attitudes that Husserl delineates.

Over the course of several decades, historians of science Giovanni Vacca (1909), W. H. Bussey (1917), Hans Freudenthal (1953), and Sabetai Unguru (1991) disagreed over who was the inventor of mathematical induction: Blaise Pascal with his *Traité du triangle arithmétique* (1665), or the Italian Francesco Maurolico with his *Arithmeticonum libri duo* (1575). While Pascal first expressed mathematical induction in the structure of Base Case and Inductive Hypothesis, Vacca and Bussey argued that Maurolico already *used* mathematical induction by repetitively applying a proposition: “This is not a very clear statement of a proof by mathematical induction but the idea is there. ... The argument from n to $n + 1$ seems to have been in

Maurolycus's mind" (Bussey 201-02). Freudenthal objected by pointing out that it is precisely Pascal's notation that reveals his full, conscious, theoretical mastery of infinity, while Maurolico's practical mastery by repetition constituted a dimmer, more intuitive, quasi-general grasp (22). Maurolico's method was an advancement from the more finite grasp of the ancients, but not on the same level as Pascal's.²³ (Unguru attacks Freudenthal's argument, but a closer analysis reveals that he misconstrues him and that the two in fact agree.) Only our underlying advanced grasp of infinity makes Maurolico's method seem so similar to Pascal's.

This debate crystallizes the curious tendency inscribed into induction itself: formalizing itself from 1) unthematized repetitive application (the ancients) to 2) *conscious* utilization of iteration (Maurolico) to 3) a mathematical formalization of that principle (Pascal). Transitioning from stage 1) to 2) becomes self-evident after one has gone through enough repetitions, but precisely when there have been "enough" cannot definitively be said. Similarly, the transition from 2) to 3) becomes evident to some at a certain point, but not to others. There is an organic, genetic trajectory from 1) to 2) to 3) with corresponding *evolutions of evidentiary structures*. Between the two ends of the spectrum, however, a fundamental attitudinal change occurs, as thinking turns from inductive to deductive. Our conception has to have completely shifted from conceiving of the world as inexact and resisting mathematical applications to perceiving it as a mathematically ideal, closed system within which infinity has become manageable. The historical process elucidates induction's inherent proclivity to formalize itself into deduction (under the name of mathematical induction). Induction technicizes itself *qua induction*. Moreover, the Pascalian moment of consciously grasping the method of induction coincides with the moment it *ceases* to be induction. The burst by which

²³ The same shift occurred later and independently between other individuals as well, for example in Jacob Bernoulli's posthumous *Ars conjectandi* (1713) of John Wallis' *Treatise of Algebra* (1685) (see Cajori).

induction comes to consciousness and shifts the evidentiary frame in the process is another instance of “Entdeckung-Verdeckung” that Husserl speaks of. Indeed, the Galilean moment *was* a moment of induction’s self-formalization into deduction.

Husserl is not content with telling a story about the methodology and worldview of science. As mentioned at the outset, Husserl’s ambition is to heal the rift between science and life. The rich notion of induction that Husserl sketched out is not merely a form of scientific inquiry, but the operation that animates life itself. As will become clear in the following pages, the lifeworld—the producer and product of lived experience—is inherently and thoroughly inductive. The lifeworld’s task is to ground, explain, and make possible the shift from the natural (inductive) to the theoretical (deductive) attitude. Its task is to keep the two conceptually distinct, while providing a genealogical link between them.

II. THE INDUCTIVITY OF THE LIFEWORLD—AND OF HUSSERL’S THINKING

1. The Lifeworld in Section 9: Anticipation, Horizon, Uniform Causal Style

Husserl draws a strong connection between life and induction when he first introduces the “Lebenswelt als vergessenes Sinnesfundament der Naturwissenschaften” (*Krisis* 48) at the end of section 9. “Was leisten wir durch [die Galileische Kunst] wirklich? Eben eine ins Unendliche erweiterte *Voraussicht*. Auf *Voraussicht*, wir können sagen, *auf Induktion beruht alles Leben*.” (51) Husserl therefore concludes his Galileo section with the radical pronouncement: *to live is to induce*. He inscribes induction into life and primitive experience itself. Husserl’s key move at the heart of the Galileo section—collapsing life and induction—is significant for two reasons. First, this account, if successful, points to a clear path to Husserl’s therapeutic goal of healing the crisis of the European sciences, i.e. the latter’s separation from life. Husserl inscribes a distinctive logic into life, a logic that is not a rote, rationalist deductivism, but an

open, self-correcting, living process. Second, the lifeworld is, under this perspective, an operative concept that guarantees the inductivity of life. If the truth that to live is to induce has been forgotten and is retrieved by a *Rückbesinnung* to the lifeworld (as Husserl performs it in subsection 9 (h)), then the lifeworld serves as a reminder of life as an ineluctably inductive practice.

Husserl continues to flesh out the link between lived experience and induction by foregrounding their anticipatory character.

In primitivster Weise induziert schon die Seinsgewißheit einer jeden schlichten Erfahrung. Die “gesesehenen” Dinge sind immer schon mehr als wir von ihnen “wirklich und eigentlich” sehen. Sehen, Wahrnehmen ist wesensmäßig ein Selbsthaben ineins mit Vor-haben, Vor-meinen. Alle Praxis mit ihren Vorhaben impliziert Induktionen. (51)

Experience is inductive in that it reaches beyond itself, i.e. beyond what it immediately apperceives. Primitive perception is already active and discriminatory for Husserl. Husserl makes this claim even more forcefully in his 1935 *Erfahrung und Urteil* (released posthumously in 1939), an elaborate study in the genealogy of logic and the emergence of formal universal judgments out of their living, pre-predicative, pre-formalized undercurrents.²⁴ In the early pages, he unambiguously inscribes inductivity into the core of experience’s horizon structure:

“Horizont” bedeutet hierbei die wesensmäßig zu jeder Erfahrung gehörende und von ihr untrennbare *Induktion* in jeder Erfahrung selbst. Das Wort [Induktion] ist nützlich, da es vordeutet (selbst eine “Induktion”) auf die Induktion im gewöhnlichen Sinne einer Schlußweise und darauf, daß diese letztendlich bei ihrer wirklich verstehenden Aufklärung zurückführt auf die originale und ursprüngliche Antizipation. (*EU* 28, emphasis in the original)

Just as in section 9 (h) of the *Krisis*, this passage draws a deep connection between induction and anticipation. This “ursprüngliche Antizipation” is an inextricable trait of experience.

²⁴ Indeed, if the *Krisis* traces the emergence of natural, mathematical sciences out of the pre-conceptual lifeworld, then *Erfahrung und Urteil* performs an analogous inquiry for formal logic. Predicative judgments are already anticipated in in perceptive (and as such discriminatory) experience.

However, induction, for Husserl, is not a just forward-looking operation of perception towards anticipated experiences beyond a given sense-perception, but also a backward-looking one. The entire passage quoted above is introduced in a section entitled “Die Horizontstruktur der Erfahrung; typische Vorbekanntheit jedes einzelnen Gegenstandes der Erfahrung.” In other words, the future anticipations of forward-looking induction are made based on a pre-known, pre-understood world. It is only because there is a past store of experiences, contributing to a given typicality that a certain range of future possibilities presents itself. The unifying frame for this dynamic is the phenomenological concept of horizon. Expected actions arise in light of a whole horizontal structure that has been experienced. “‘Horizont’ bedeutet hierbei die wesensmäßig zu jeder Erfahrung gehörende und von ihr untrennbare *Induktion* in jeder Erfahrung selbst” (*EU* 28). This horizon can be corrected in light of new, unexpected experiences that are incorporated into what is known. Inductive experience is always implicated into a dynamic of zig-zagging back and forth between past and future. Induction is a time-problematizing operation and as such inexorably historical.

An upshot of such a constantly zigzagging practice is that the world is endowed with a transtemporal, uniform causal style. Perceptive experience’s constant anticipation of what is yet unknown and to come and its frequent checking back against, and correction of, the past result in a stable, unified world that is presumed to persist through time. This practice is iterative and self-referential, i.e. it repeatedly corrects and amends itself to provide this stability. The lifeworld is both the substratum and result of this practice. Inductive experience constantly acts in it, modifies it with its transcending sense of foresight, and, interestingly, thereby sustains security and constancy. This is what Husserl means when he speaks of the “Bewährungscharakter” of experience in the context of his exposition of the lifeworld in section 34 (d) of the *Krisis* (138, Hua VI, 130). The lifeworld is always already familiar, yet

constantly evolving and technicizing itself. It is the corollary to Husserl's operative term of inductive experience.

Such a stabilized world can only arise out of a time-problematizing experiential operation because this operation is inherently self-referential and iterative: in a cybernetic feedback cycle, induction repeatedly refers back to itself as it anticipates and takes in new data, and confirms or corrects past anticipations based on those new data. An inherently historical practice tries to overcome contingency by fostering a reliable, corroborated, functional habituality. Induction (properly executed) is constantly on its guard, in touch with the world, and protective of life by developing grounded, stable structures.

Inductivity is thus an operation that is at the very heart of several of his key phenomenological concepts, i.e. the lifeworld, experience, and the horizon. The inductive experiential practice of life that moves from a pre-understood past to new open possibilities is thus central to understanding Husserl's concept of the horizon. This applies to the internal horizon on the one hand, which refers to an ever closer and more approximate explication of an individual object ($E\mathscr{C}U$, §8), and to the external horizon on the other, i.e. the background of other objects to which the object that is consciously thematized refers in a murky, still unexplicated manner ($E\mathscr{C}U$, §8). Indeed, induction, in this expansive Husserlian sense of a dynamic, temporal, living logic, is inscribed into phenomenology's most central concept of all: intentionality. Intentionality, the essential aspect of consciousness, is always, qua intentionality, directed at something. To be conscious is to intend an object (whatever its ontological status may be or is conceived to be by that intending consciousness). Intentional consciousness is thus eo ipso transcendental, in that it anticipates something for which it seeks fulfillment (Erfüllung) and confirmation (Bewährung). Consciousness is thus nothing but a cashing out of the inherently inductive tendencies of primitive experience.

2. The Humean Problem and the Reentry of the Lifeworld

After introducing the lifeworld at the end of the Galileo section, Husserl leaves it behind and does not return to it for some time. He reintroduces it in sections 33 and 34 in the *Krisis*, after a tour de force through the history of philosophy from Descartes and the subsequent rift between the English empiricists—culminating in Hume—and the European rationalists—culminating in Kant. While both lines of thought operated under a naïve view of objectivism for Husserl (i.e. the Galilean-Cartesian notion that the world is perfectly exact, measurable, and governed by mathematical laws) (*Krisis* 83-84, 93), Husserl disagrees particularly strongly with Kant. Kant's transcendental idealism was unable to properly explain the status and ground of the mathematical sciences, as it simply the apodicticity of mathematical truths as evident and given (92-95, 118). Kant overlooked the subjective, sensual realm in which the judgment of objective sciences' apodicticity is generated. This is because, Husserl argues, Kant had only a facile understanding of Hume's central problem, which is to say—the problem of induction. It is precisely at this juncture of criticizing Kant and re-engaging Hume's problem that Husserl introduces the concept of the forgotten lifeworld for a second time. He says:

Kant spricht vom “*Humeschen Problem*”. Was ist das wirkliche, das *Hume selbst* bewegende? ... Wie ist die *naive Selbstverständlichkeit* der Weltgewißheit, in der wir leben, und zwar sowohl die Gewißheit der *alltäglichen* Welt, als die der gelehrten *theoretischen Konstruktionen* aufgrund dieser alltäglichen Welt, zu einer *Verständlichkeit* zu bringen? (99)

Hume, through his problem of induction, was already hinting at a pre-understood world in which we operate, a necessary grounding concept that he failed to develop and that Kant's rationalism excluded. Just like in section 9 (h), Husserl thus introduces the lifeworld as a corrective concept to an unquestioning, misguided acceptance of deductive thinking as self-evident and not in need of further grounding. For the second time, Husserl relies on the

doctrine of the lifeworld when he argues for a hidden, pre-understood but unthematized underbelly to a seemingly self-sufficient mathematical, deductive structure that of this world. Both in the context of discussing Galileo and Kant, the lifeworld is introduced as the indispensable but forgotten ground of life out of which mathematical science and rational, deductive thinking emerge (106-07).

Husserl seeks to correct Kant's failure to engage with Hume with the concept of the lifeworld: Kant had always already supposed the lifeworld as the ground of the sciences and rational thinking, which prevented that lifeworld from coming into view. Studying the lifeworld aims to make this shift intelligible and to ensure that we remain aware of life and the lifeworld's ineluctable inductivity. Hume was wrestling and struggling with a problem—the problem of induction—that Kant, because of his rationalist deductivism, failed to see and address. The lifeworld fulfills the strategic function of anchoring in life an inductive practice that leads to formalizations, cultural achievements, and self-otherings.

The lifeworld cannot be separated from Husserl's doctrine of inductivity. As Husserl fleshes out in Part paragraph 34 (d), the lifeworld is given in direct, unquestioned, and corroborated experience:

Die Lebenswelt ist ein Reich ursprünglicher Evidenzen. Das evident Gegebene ist ... in Wahrnehmung als "es selbst" in unmittelbarer Präsenz Erfahrenenes oder in Erinnerung als es selbst Erinnertes; ... jede in diese Sphäre gehörige mittelbare Erkenntnis, weit gesprochen: jede Weise der Induktion hat den Sinn einer Induktion von Anschaubarem. ... Auf diese Modi der Evidenzen führt alle erdenkliche Bewährung zurück. (130)

This realm is the practical, kinesthetic, subjective, and intersubjective context in which we operate, an always unified but vague horizon in which interests, anticipations, and goals present themselves to us (111). It is not surprising to find the term "Induktion" crop up again at this point in the *Krisis*. The lifeworld provides the originary evidences that trigger inductive practice and catalyze it. The lifeworld is thus not just the primitive, lost world that we inhabited

before there was science. Before the advent of science, our lifeworld may happen to be pre-scientific; however, when the theoretical attitude takes over, we still operate in familiar, pre-understood evidentiary structures, only this time with scientific knowledge and assumptions woven into our perception of the world. With the advent of science, we do not leave the lifeworld, but begin to inhabit a different one. The lifeworld is only “pre-scientific” insofar that scientific thought and practice are grounded in direct experiences and modes of engagement for which science itself cannot account.

By the same token, the lifeworld is also not the simple aggregate of all objects. Indeed, the lifeworld is inherently unobjectifiable: it is the ground on which objectification occurs (132). In short, the lifeworld enshrines the continued unfolding of life. In his exposition of the lifeworld, Husserl states:

Die Seinsbewährung des Lebens ergibt in Erfahrung terminierend eine volle Überzeugung. Selbst wenn sie induktiv ist, ist die induktive Antizipation die einer möglichen Erfahrbarkeit, die letztlich entscheidet. Induktionen können sich durch Induktionen im Miteinander bewähren. ... [D]a jede direkte Wahrnehmung selbst schon induktive Momente ... einschließt, so ist alles im weiteren Begriff “Erfahrung” oder “Induktion” beschlossen. (n.130)

The lifeworld is the indispensable substratum that anchors the practice of inductive life and keeps it running: anticipatory, interested, self-clarifying, and self-formalizing. As this operation is inherently time-problematizing and trying to counter the flow of time with an abiding stability that it derives from the experiences it accumulates, the lifeworld itself is thus a sedimented, historical product that abides in its essence, yet evolved in its specificity. The lifeworld is the substratum of inductive life by being both the basis on which it unfolds and the product of that practice, which is then channeled anew into living experience.

Forgetting the lifeworld’s inductivity when trying to elucidate it and succumbing to objectifying it in a transparent, systematic account is thus the cardinal sin that Husserl tries to avoid and warn against (133). He does not want to provide an objectivist analogue to the

scientific concept of “world” (135-136), but instead asserts repeatedly that he is interested in the “Seinsweise” of the lifeworld (126, 128, 129, 134) and the “Wie” of its givenness (146, 147, 161, 163, 170). He is aiming for a “Wissenschaft von dem universalen Wie der Vorgegebenheit der Welt, also von dem, was ihr universales Bodensein für jedwede Objektivität ausmacht” (148). A full appreciation of the inductivity of life, and the lifeworld’s maintenance of it, rules out the objectivist misunderstanding of a systematic, fully transparent account of the lifeworld. For if the lifeworld enshrines a constant continuous unfolding of inductive life, and if induction is a necessarily self-clarifying and “entdeckend-verdeckende” practice, then making the structures of the lifeworld transparent would bring life to a halt: having fully uncovered the conditions that determine life, its interests and dim anticipations would also become fully apparent, foreclosing all further possibility for life to clarify itself. Life would cease to be inductive once it has been locked into a systematic frame.

3. The Motivational Problem

With this account of the lifeworld’s inductivity in place, it is worth returning to the scholarly debates that criticize Husserl’s theory of the lifeworld as irreconcilably torn between an ontological and transcendental account. It should be clear now that the scholarly debate has 1) framed the question under misleading terms and 2) overlooked the answer Husserl attempted to provide. The dichotomy “ontological vs. transcendental” is not something that Husserl overlooked, but sought to reconcile with his rich account of induction. As the inductive practice of life, by anticipation, always already reaches beyond that which it perceives, and as the anticipated ideal world influences how and what is experienced in the first place, the lifeworld must necessarily be understood as both ontological and transcendental. Induction—more precisely, inducing—is only conceivable as at once ontological and transcendental.

Husserl himself was not exempt from toying with a deductive, systematic account of the lifeworld, as he admits toward the end of Part III.A: “Treiben wir nicht auch Wissenschaft, stellen wir nicht Wahrheiten fest über wahres Sein? Kommen wir nicht auf die gefährliche Bahn der doppelten Wahrheit?” (179). However, it is important to note that he makes the attempt, “den zum Nachverstehen Willigen einen der Wege zu führen, die ich wirklich gegangen bin” (123), a path of thinking which, as it attempts to develop a new method among uncharted territory, is subjected to self-correction and self-clarification as it reflects and looks back on itself. Husserl does not stick with his objectivist conception of the lifeworld, but moves beyond it in the face of appearing paradoxes (§§52-54). Husserl’s last section in Part III.A is therefore entitled “Die prinzipielle Korrektur unseres ersten Ansatzes der Epoché [. . .]” (190). He reiterates:

Es ist natürlich ein lächerliches ... Missverständnis, die transzendente Phänomenologie als ‚Cartesianismus‘ bekämpfen zu wollen, als ob ihr ‚ego cogito‘ eine Prämisse oder Prämissensphäre wäre, um aus ihr die übrigen Erkenntnisse ... in absoluter ‚Sicherung‘ zu deduzieren. Es gilt nicht, Objektivität zu sichern, sondern sie zu verstehen. Man muss endlich einsehen, dass keine noch so exakte objektive Wissenschaft irgend etwas ernstlich erklärt oder je erklären kann. Deduzieren ist nicht Erklären. ... Das einzig wirkliche Erklären ist: transzendental verständlich machen. (193)

To avoid such deductivism, Husserl proposes a new attempt at finding a path into phenomenology, one that builds on the previously recognized dangers of misunderstanding the lifeworld and its associated epoché. This time, he begins not from the lifeworld, but from subjectivity itself.

It is worth pausing over Husserl’s wavering and more repetitive style in Part III.B of the *Krisis*. Time and again he circles back to the danger of an objectivism, the transcendental epoché, how it will overturn our whole way of thinking and experiencing, its proper understanding, and the importance of its radical execution. Husserl’s style of thinking clearly

aims at an extremely purified notion of the epoché, which he hopes to reach by a stepwise, iterative path of understanding:

Die phänomenologische Psychologie erschließt sich ihrem Sinne nach in verschiedenen Stufen, weil die phänomenologische Reduktion selbst – und das liegt in ihrem Wesen – ihren Sinn ... nur in Stufen erschließen konnte. (250)

Indeed, his traversing of the scientific (Cartesian) epoché in Part II, the epoché of the lifeworld in Part III.A, and the transcendental epoché in Part III.B are the three major stepping stones on that path. Both the goal and the method become clearer and clearer along the way, and they aim for a sudden, radical rewiring of our thinking in a shifted evidentiary structure: once the transcendental epoché has truly been performed, the phenomenological mode of thinking will become evident (251). True to life, Husserl's own thinking is inductive, trying to guard against the trappings of deduction.

More precisely, yet, Husserl attempts to utilize the inductivity of living thinking and its proclivity to formalize itself and alter its evidentiary structure to usher in a shift of attitude. Husserl wants to perform a Galilean moment, only this time with a “*vollbewußt verstandenen und geübte Methode*” (46), i.e. in a radically new way. It is an attempt that Husserl does not just theorize but practices—inductively.

If Blumenberg thus rules out the possibility that the theory of the lifeworld could explain the *Umstellung der Einstellung*, he is off the mark. The historically contingent occurrence of such a shift does *not* imply that there is no explicable basis for this shift to happen. Husserl's theory seeks to make the shift plausible by inscribing it into the very dynamic of inductive experience. Yet, it has to be conceded that his theory does not go very far in making intelligible the *moment* in which such a shift is triggered. What does it mean or look like for such a shift to occur and unfold? Is such a moment even accessible to our

understanding?²⁵ This is where Husserl's preliminary, not fully developed theory of induction abandons us. The moment of transition, as, for example, in the Galileo section, is identified, stated, and analyzed for its implications, but the moment itself is never elucidated. Its possibility is argued convincingly, but its actualization is not shown. This moment of the *Umstellung*—the leap in knowledge—deserves more excavation.

III. HUSSERL AND HUME'S PROBLEM OF INDUCTION

1. The Humean Account

We thus have to look to similarly minded philosophical allies of Husserl's to fill this gap. As discussed above, Husserl saw Hume as wrestling with a problem—the problem of induction—that Kant, because of his rationalist deductivism, failed to fully grasp and address. Since Husserl's account leaves the transitional moment from induction to deduction unclear, it is thus useful to look to Hume's account of induction, and see how he grapples with the problem of grounding deduction. From there, this chapter will draw on an unlikely source to fill this gap: Aristotle.

At first blush, it might be surprising to cast Husserl, a trained mathematician and long-time adherent to Neo-Kantian thought, as someone who continued Hume's intellectual project. But some scholars have already traced the substantial affinities between Hume and Husserl. Richard Murphy argues that Husserl's philosophical maturation took several important cues from Hume, such as the idea that objectivity and rationality emerge out of subjectivity, i.e. lived acts of consciousness (Murphy 1979, 203). Because of this Humean approach, Husserl's notion of the *a priori* was much less formalistic than Kant's (Murphy 1980,

²⁵ This is the elusive moment of transition that this chapter has been zeroing in on since its initial investigation of the elusive shift of philosophical induction to mathematical induction, i.e. from induction to deduction.

92), which lead to the “distinctive character of Husserl’s transcendental idealism. It is not Kantian for it contains empiricist strains totally alien to Kant’s outlook” (Murphy 1980, 104). Indeed, as early as 1908, Husserl claimed that “‘Humean psychology is the first systematic endeavor of a science of the pure givenness of consciousness ... Hume’s *Treatise* is the first draft of pure phenomenology’” (Murphy 1979, 213).

It is thus no surprise to see those empiricist strains become especially pronounced several decades later in the *Krisis*. More specifically, Mall has observed that Hume and Husserl fundamental agreement that experience, practice, human life as such, are thoroughly inductive: “Induktion ist nach Hume eine Operation der menschlichen Natur” (Mall 45). Indeed, both for Hume and Husserl “Die Sinne scheinen selbst induktiv zu arbeiten ... So ist jede Wahrnehmung antizipierend. ... Das Erwartungsmäßige liegt schon jeder Induktion zugrunde” (Mall 51). It is thus warranted to briefly lay out Hume’s account of induction and experience’s inductivity to clarify the problem that Husserl sought to solve: the transition from inductivity to deductivity.

Hume divides the cognitive states (“perceptions,” as he calls them) into two categories—sense impressions and ideas. The latter derive solely from the former (*Treatise* § 1.I.i., p.2-3; *Enquiries* § 12, p.18).²⁶ Ideas can be connected or associated with each other through the relations of resemblance, contiguity, or causation *Treatise* § 1.I.iv., p.11; *Enquiries* § 19, p.24). The relation that captures much of Hume’s thinking in the *Treatise* and the *Enquiries* is causation. We infer the notion of causation, i.e. of a universal dictating a necessary, non-contingent connection between events, from a frequent succession and contiguity of those events (*Treatise* § 1.III.ii., p.75-76; *Enquiries* § 22, p.26-27). But since this inference is derived

²⁶ The argument is laid out in *A Treatise of Human Nature* and reappears, with minor modification, in *Enquiries Concerning Human Understanding*.

from limited and contingent empirical sense data, Hume wonders how we can ever make an inference on the universal necessity of that connection (*Treatise* § 1.III.iii., p.79; *Enquiries* § 28-29, p.32-34).

Hume explains that it is the constant conjunction of two events in our sense impressions that forces the inference of necessary connection upon us (*Treatise* § 1.III.vi., p.87). Yet, this inference can be justified neither demonstratively (deductively) nor probabilistically (inductively, or from the cases under examination), for in both lines of reasoning, we are assuming that our past experiences have purchase on future events. On a deductive justification, Hume concludes: “Our foregoing method of reasoning will easily convince us, that there can be no *demonstrative* arguments to prove, *that those instances, of which we have had no experience, resemble those, of which we have had experience* (*Treatise* § 1.III.vi., p.89, emphasis in the original; *Enquiries* § 32, p.37). Similarly, inductive, or probabilistic reasoning falls short on account of a circularity: “The idea of cause and effect is deriv’d from *experience*, which informs us, that such particular objects, in all past instances, have been constantly conjoin’d with each other. ... According to this account ... probability is founded on the presumption of a resemblance betwixt those objects, of which we have had experience, and those, of which we have had none; and therefore ‘tis impossible this presumption can arise from probability” (*Treatise* § 1.III.vi., p.89-90, emphasis in the original; see also *Enquiries* § 32, p.37-38).

Since it is precisely the assumption of a continuity of nature from past to future that is in need of justification when it comes to arguing for a necessary (i.e. causal) connection between events, induction turns out to be an ungrounded inference for Hume. Falling short of finding such a justification, Hume can only explain how human thinking draws this inference to a universal law without such a justification: It relies on *habit* and *custom*. “Objects have no discoverable connexion together; nor is it from any other principle but custom

operating upon the imagination, that we can draw any inference from the appearance of one existence to the existence of another” (*Treatise* § 1.III.viii., p.103; see also *Enquiries* § 32-33, p.37-39). Hume identifies this mode of reasoning as an innate instinct and an inexorable, yet unknown operation of the mind. “Twill here be worth our observation, that the past experience, on which all our judgments concerning cause and effect depend, may operate on our mind in such an insensible manner as never to be taken notice of, and may even in some measure be unknown to us” (*Treatise* § 1.III.viii., p.103; see also *Enquiries* § 38, p.46-47). Hume further states that

The custom operates before we have time for reflection. ... [E]xperience may produce a belief and a judgment of causes and effects by a secret operation, without being once thought of. This removes all pretext ... for asserting that the mind is convinc’d by reasoning of that principle, *that instances of which we have no experience, must necessarily resemble those, of which we have.* (*Treatise* § 1.III.viii., p.104; see also *Enquiries* § 36, p.43-45)

Hume’s problem of induction thus traces an analogous path as Husserl’s: from a repetition of certain sense impressions, a pattern of lawfulness gradually establishes itself that makes it inevitable for human thinking to expect and anticipate that events in the surrounding world will continue to follow the patterns previously observed. (The only difference seems to be that Hume derives of this epistemological development from psychological operations, while Husserl inscribes it into the much broader and vaguer notion of life.) Hume ran into a dead end when he stumbled upon the lifeworld, which he never brought under a name or concept, but could only conceive under the aegis of a problem. Deductive reasoning only becomes possible on the presupposition of such an (unjustified) inductive basis.

2. Husserl’s Response to Hume’s Problem of Induction

In his supplements to his discussion of empiricism in paragraphs 23-28 of the *Krisis*, Husserl describes how British empiricists generally, and Hume in particular, were on the cusp of grasping the solution to how objectivity emerged out of subjective experience:

Es liegt in der Einstellung des Empiristen, daß er die wirklich anschauliche, wirklich erfahrene und erfahrbare Umwelt zum Ausgang und Thema nimmt und danach fragt, wie von ihr aus das Denken, die *operations of mind*, über das Erfahrene und Erfahrbare hinausführen sollen, im besonderen als Methodik der Wissenschaften *more geometrico*. (Hua VI, 448)

Yet, the empiricists' inability to shed their psychologistic, naturalistic presumptions about living acts of consciousness--ascribing them to the same order of being as material entities (Hua VI, 433, 450)—removed the possibility of transcendence and confined empiricists to a constant battle with the threat of solipsism (Hua VI, 432, 448).

Husserl's congruence with Hume with regard to the inductive operations of the mind, as Mall has worked out in detail, is thus not merely incidental, but rooted in the fact that Husserl wanted to revive Hume's attempt to realize the Cartesian project of laying out the structures of subjectivity. Those structures are not those of an a priori deductivism, but thoroughly inductive. There is not merely an analogy between Husserl's problems and the problem of induction, as noted earlier. Rather, Husserl's motivational problem is a reappearance of Hume's problem of induction, precisely because Husserl takes up his philosophy again. The lifeworld's strategic task is to fill that gap, to solve the problem that Hume laid out. It does this by turning Hume's problem—and acknowledgment that human experience infers an inductive causal unity to the world even though there is no justification for it—into an explanatory concept: Given that we do link inductive experience to a deductive, predictable uniformity of the world, there must be some concept that ensures their connection, even if it operates unnoticed. Husserl's lifeworld is the missing concept that responds to the problem of induction by thinking induction and deduction together.²⁷

²⁷ We need to live in a uniform, predictable world, for otherwise there can be no living at all. (For if there is nothing familiar, habitual, and pre-understood, how are we to act and live?)

Hume therefore presented the motivational question to Husserl, but left him to his own devices when it came to addressing it. The crucial difference between Husserl's and Hume's notions of induction is, of course, that Hume's version does not provide access to ideal objects. Objects of mathematics and geometry belong to the knowledge realm of relations of ideas, which are by definition separate from the knowledge realm of matters of fact, which ground experiential, habitual thinking. Hume's problem of induction precisely boils down to this: somehow, the mind makes the transition from observed regularities to relations of ideas, but we cannot give a justification for this transition. Husserl's revival of Hume's thought elucidates why this gap becomes unbridgeable to Hume (i.e. because of the sensualist naturalization of lived acts consciousness that rules out any possibility of transcendence). Hume severs the empirical and the ideal, matters of fact and relations of ideas, the inductive and the deductive. It becomes clear that Husserl's lifeworld has to fulfill the strategic function of avoiding this mistake and bridging the gap. Hume, on the other hand, does little to elucidate Husserl. The problem of induction reappears as the motivational problem, and it is incumbent on Husserl's lifeworld and the inductive operation that it implies to address it.

Husserl begins to respond to the Humean problem by addressing both questions at once with his concept of the lifeworld: induction and deduction cannot be separate, for the universal law is that which is anticipated and induced toward; without such an anticipation, there could be no induction in the first place. Conversely, every deductive calculus that emerges out of induction is fed back into that inductive practice and is used inductively. For Husserl, ideal objects and deductive thinking are not only an outcome of self-referential induction, but inductive practice would never begin in the first place for Husserl if it was not already aiming at an ultimate (albeit unclearly and dimly anticipated) ideal pole beyond the real.

Induction and deduction are not only interlinked, but incoherent if they are thought as separate from each other. What remains to be conceptualized is *how* they ought to be thought together.

IV. THE MISSING LINK: ARISTOTELIAN NOUS AND EPAGOGÉ

1. The Role of *Nous* in Grasping the Universal from Repeated Sense Perception

The question is where to look for the missing link. One place would be theories from Husserl's time that sought to solve Hume's problem—such as the account devised by Carnap (see Carnap a, Carnap b, Nagel). But a survey of the theories of the time suggests that nearly none of those accounts hold the answer, as they sought to subsume induction under deduction (and thus never address the question of how induction turns into deduction), declared the two incompatible, or left their hierarchy unspecified (Brody). To find the missing piece in Husserl's theory, we instead have to go back—perform a *Rückbesinnung*—to a theory that does not approach induction and deduction as a rifted dyad that has to be reconciled. The prime candidate for this theory can be found with a thinker against whom Hume had positioned himself: Aristotle and his notions of nous and epagoge.

Indeed, upon closer investigation, it will become clear that, consciously or not, Husserl is retrieving exactly this oldest of theories of induction and its underlying thought. Aristotle and Husserl share extremely similar basic assumptions and aspirations for what their accounts of induction are meant to achieve. For both, deduction stands in for rigorous, incontrovertible, universal, and trans-temporal knowledge that is deductive and scientific. For both thinkers, deduction builds up the fully systematized body of knowledge that is true and apodictic. It makes up the realm of episteme, while pre-deductive knowledge is characterized as doxa and is derived by induction. It is a practice that is much less formalized, but has a tendency to

formalize itself and lead to the deductive kind of thinking that is employed in the arts and sciences. Both thinkers therefore provide a genealogy of deduction out of induction, with striking similarities on the level of overall structure, as well as small details.

Such a claim is not easily argued, as Aristotle's remarks on induction are scattered throughout his works, in sharp contrast to his systematized theory of deduction, i.e. his work on the syllogism in the *Prior Analytics*. There has been much debate about how Aristotle's theory of is to be understood, whether or not it is coherent, and what role it plays in his philosophy as a whole (Groarke 4-8). Wayne Thompson, in *Aristotle's Deduction and Induction*, acknowledges that "the common belief that his induction is incomplete, unsystematic, and generally unsatisfactory as compared with his deduction is well justified. [Yet,] induction still is vital to the total system of scientific thought" (Thompson 81). Nonetheless, there exist useful summaries and systematizations of Aristotle's doctrine of induction. Watson provides a clear rendering of his major steps. Thompson himself defies the difficulties he diagnoses and provides a typology of Aristotle's different accounts of induction. Groarke, too, provides a systematic rendering of Aristotle's accounts of induction as a pre-discursive moment of direct insight into the essence of things, the subsequent formal reasoning developed on the basis of that insight, the rhetorical rendering of such thinking (well or badly executed) in rhetorical arguments, and the statistical approach of enumeration by cases (Groarke 158, 207).

A scholarly consensus has emerged on several points.²⁸ First, it is undisputed that induction constitutes a transition from particular to universal, while deduction delineates the transition from universal to particular; as such, the two are mutually exclusive. Second, the

²⁸ This was gathered from a parsing of the introductory works by Watson (1904), Barnes (1993), Thompson (1970), and Groarke (2009). What has been particularly helpful was the debate over Aristotle's theory of induction in *Phronesis* that unfolded between Heß (1970), Hamlyn (1976), Engberg-Pedersen (1979), and Upton (1981).

role of induction is to precede and ground deduction. Despite disagreement on whether the genealogy is psychological (Engberg-Pedersen) or epistemological (Bayer) in nature, there is no doubt as to the primacy of induction over deduction in the Aristotelian account. As such, the genealogy involves a transition from perception (i.e. the empirical) to essential insight (i.e. the ideal). This genesis is sketched out in *Posterior Analytics* II.19. Third, induction is a mode of insight that is neither a Platonic recollection of truths once impressed on the soul but later forgotten; nor is it a derivation of knowledge from previous knowledge, i.e. a mode of logical inference. Fourth, in the process of induction, the capacity of comprehension, or nous, intervenes to aid the jump to insight into the most basic and essential principles. The nature of this moment, and of nous more generally, is one of the most hotly contested points in Aristotle scholarship and will be taken up again below.

A survey of existing studies on Aristotelian epagoge reveals that, among the scattered remarks, there is a privileged passage to query for a deeper understanding: the last chapter of the *Posterior Analytics*.²⁹ This is not just because Aristotle devotes the entire chapter to fleshing out his conception of epagoge, but also because of the chapter's strategic location within the *Posterior Analytics* and the *Organon* as a whole: after having laid out the formal structure of the syllogism in the *Prior Analytics* and having considered its material conditions and restrictions in the *Posterior Analytics*, the work has to grapple with the question: while deduction provides us with scientific knowledge (episteme) from premises that are agreed upon as necessary and true in the confines of the specific syllogisms under consideration, whence do that truth and

²⁹ There is another, much-discussed passage in Aristotle's works that deals with induction, namely *Prior Analytics* II.23, i.e. the passage on the inductive syllogism. This passage, however, has received so much attention mostly because it is 1) at odds with the rest of Aristotle's remarks on epagoge, and 2) it gives fodder to the hope that induction under Aristotle may, after all, be able to be shown to be deductively valid and produce demonstrative knowledge. Both of those peculiarities, as well as *Prior Analytics* II.23's appearance in the context of a treatise on *deductive* logic, merely go to show that that a much simpler point obtains about the passage: it speaks about a form of deductive reasoning, not inductive reasoning.

necessity originate? What guarantees their truth and necessity? Aristotle devotes the *Posterior Analytics* to finding an answer to this question, which, in the penultimate chapter II.19 turns out to be: induction. Every true and necessary conclusion refers back to more primary premises that are necessary, truer, and better known than their conclusions. To avoid an infinite regress or a *circulus vitiosus*, there must be an ultimate ground of knowledge that is not itself deductive or deductively derived. Since all knowledge either stems from deduction or induction, the legitimacy of deductive (scientific) knowledge can only be guaranteed by induction. *Posterior Analytics* II.19 is thus worth considering as a privileged passage and an authoritative pronouncement by Aristotle on his theory of induction.

From the outset of looking at *Posterior Analytics* II.19, it is striking to see that Aristotle, like Husserl, “conclude[s] that these states of knowledge are neither innate in a determinate form, nor developed from other higher states of knowledge, but from sense perception” (*APo* 100a9-11). Aristotle pushes this point in opposition to the doctrines of anamnesis and learning by deriving knowledge from previous knowledge, as he judges both of them to be inadequate when it comes to explaining our knowledge of first and ultimate principles (*APo* 99b26-32). Instead, it is sense perception that leads to such insight. This is possible because sense perception, as Aristotle is keen to underscore, is not merely passive and receptive, but an active and discriminative faculty: humans and animals alike “have a connate discriminatory capacity, which is called perception” (*APo* 99b35-36). The next crucial step in Aristotle’s genealogy is that all animals have the capacity to retain sense impressions (99b36-37), especially when “such persisten[t impressions] is frequently repeated” (100a2). Persisting or recurring sense perceptions are systematized into a memory, which in turn gives rise to what Aristotle calls experience (100a3-5). Experience, in other words, is a unified habitual frame of engaging with the world, based on and structured by a background of memories. The past influences how

we interact with the world in the present: past occurrences and their patterns establish a horizon of expectations. From perception to repetition to retention to memory to experience, Aristotle builds up a genealogy of rational thinking that strongly resembles Husserl's in *Erfahrung und Urteil* and the Galileo paragraph.

The “again and again” of sense perception is the central feature that structures human experience and leads to the establishment of a universal: at a certain moment, a universal principle is apprehended from the particulars perceived (but the universal is not itself perceived in the same way as particulars are). This apprehension occurs with the crucial help of nous (an intuitive ability to directly grasp and comprehend), which, when it kicks in, makes the perceived particulars appear in a different light: the universal, which was anticipated in the particulars without being identified as such, reveals itself with clarity (100a15-17). The principles thus apprehended provide the starting point for deductive inference and the derivation of scientific knowledge. It is this process in its entirety that Aristotle calls induction (100b4).

Aristotle and Husserl thus lay out fundamentally congruent processes of induction: perception, repetition, memory, and a habitualized anticipation guide a self-technicizing process. At a certain point, this anticipatory process culminates in a jump in which universal principles are grasped and locked in to serve as grounds for deductive thinking: a rational structure has emerged. Just like Husserl, Aristotle inscribes this inductive process into active, discriminatory faculties of sense perception and retention, i.e. the living practice of any human (*APo* II.19; see also *Metaphysics* A.1). *Science is rooted in life for Aristotle, as it is for Husserl.* Just like Husserl's, Aristotle's genealogy of deduction from induction draws a connection from the real to the ideal, from the empirical to the universal, and from the ontological to the

transcendental. Aristotle insists that “the soul is such as to be capable of undergoing this” (*APo* 100a13-14) transition from finite empirical knowledge to a grasping of the universal.³⁰

Indeed, scholarly disagreement and debate emerged over the crucial moment in the inductive process: the intervention of nous by means of which the particular is transcended and the universal principle apprehended. The moment of nous breaking through is when the hitching of the inductive to the deductive occurs and transition from the former to the latter is made. Watson, for example, in his summary of the process, vacillates between acknowledging Aristotle’s lack of justification for that transition and questioning whether this moment of nous seizing on the principles is so obvious that it need not be justified in the first place:

Though it can hardly be denied that the transition from invariable concomitance [of two observed attributes] to absolute invariability is hard to justify, it must be said, in defence of Aristotle, that his doctrine is based upon the principle that nature is not a sphere in which pure contingency prevails, but is on the whole subject to law. This, indeed, is a presupposition for which Aristotle can supply no justification; but, granting its truth, it is natural to suppose that when by induction we have discovered certain invariable conjunctions, the mind is able to seize upon the universal principle which these conjunctions suggest. (Watson 148).

In an article for *Phronesis* in 1970, Walter Heß saw the transition from observed correlations to a grasp of the underlying essence, or *eidos*, as more problematical and sparked a debate over what exactly this intuitive, apparently non-methodical element in induction had to be understood (Heß). Hamlyn responded by arguing that Aristotle does not show the method by which nous moves from particular, inductive observations to universal, deductive principles

³⁰ Husserl’s close epistemological, logical, and psychological affinity with Aristotle should thus challenge the common perception that Husserl is a secret Platonist. Blumenberg, for example, charged Husserl’s lifeworld theory and its epistemology with reviving the old account of Platonic anamnesis in “Das Lebensweltmißverständnis.” Indeed, if my account of Husserl’s affinity with Aristotle is correct, the Platonist charge could not be further from the truth: Husserl espouses a view that is in direct opposition to the Platonic epistemology of recollection.

(Hamlyn). Engberg-Pedersen retorted that a grasping faculty like nous can eo ipso not operate by drawing inferences, as that would already consign it to the realm of deduction (Engberg-Pedersen). Nous was simply an intuitive capacity that could seize on the universals directly, although it was liable to the risk of making errors: nous can seize on universals that turn out not to be universals: “what I mean by nous is something much humbler [than others have suggested]: a generalizing capacity that is responsible for the fact that a universal point, something that is, which goes beyond sense perception, may come to be present to the mind – whether this point be true or false” (Engberg-Pedersen 308). Upton subsequently took Engberg-Pedersen to task for this latter point and insisted that nous, as a capacity that intuitively seizes on universal principles, could not be deceived in that process (Upton 175). This brought the discussion back to square one: nous is said to be able to seize on the universal principle directly, yet it is unclear *how* it does that. Barnes, in his extensive commentary to his own translation of *Posterior Analytics*, tries to make a first step towards alleviating the problem by translating nous not as the mystical term “intuition,” but the more pedestrian “comprehension” (Barnes).

It is thus still unclear how nous, and with it the moment of transition from particular to universal principle, has to be understood. However, scholarship does largely agree that “Aristotle sees no Humean problem about a leap from inductive evidence to *knowledge* ... Hence, as *he* sees the problem of our grasp of first principles, the difficulty is not a lack of evidence to transform inductive belief into certain knowledge” (Burnyeat 131). While some maintain “dass bereits Aristoteles das Problem der Induktion in Gestalt der Frage kannte, was uns denn angesichts der Unmöglichkeit einer alles Einzelne umfassenden Beobachtung jeweils berechtige, von einer bestimmten Beobachtung in n Fällen auf die Möglichkeit der nämlichen Beobachtung im Falle $n + 1$ zu schließen,” (Heß 60), they do concede that Aristotle seemed

confident to adequately address and solve that problem. Indeed, some argue that the problem of induction emerged from a misconstrual of the Aristotelian account on Hume's part: "Some commentators leave us with the impression that Hume understood the Aristotelian and medieval points of view on induction and refuted them. It is more plausible to suggest that he never seriously considered these doctrines in the first place" (Groarke 70).

2. Epagoge, the Analogy of the Rout, and Exemplarity

If it is agreed, then, that induction does *not* pose a problem to Aristotle (for it clearly, in fact, occurs), then neither does the motivational problem of the attitudinal change. How then, does Aristotle make the transition understandable? Aristotle illustrates the problematic moment of insight into the universal principle in the notorious analogy of an army regrouping after a rout. "The [primitive] states neither inhere in us in a determinate form nor come about from other states which are more cognitive; rather, they come about from perception—as in a battle when a rout has occurred, first one man makes a stand, then another, and then another, until a position of strength is reached. And the soul is such as to be capable of undergoing this" (*APo* 100a10-14; emphasis mine). It is worth focusing on the short analogy and to develop its implications and strategic deployment. A formation has been broken up and is in retreat. It then recoups itself by one individual soldier taking a stand, animating another to do the same, and so on, until a stable formation is restored. The key moment in this metaphor is the *first* soldier's sudden stop. The arrest of the disorderly retreating from overpowering and debilitating forces happens in a turn of fortune. However, it does not happen all at once and conclusively. The single soldier, perhaps instinctively, perhaps consciously, takes a stand in defiance of all the odds against him and puts his own life on the line. The stable formation need not happen, and it only does because other soldiers follow the first soldier's example of showing courage at a time of desperation. Aristotle calls this is the trajectory from perceived particulars to the

universal induction: “When one of the undifferentiated items [in perception] makes a stand, there is a primitive universal in the soul; for although you perceive particulars, perception is of universals ... Thus it is plain that we must get to know the primitives by induction; for this is the way in which perception instills universals” (100a16-b5).

The first soldier’s courage is *exemplary*: others follow his lead and stop as well, “then another does and another, until a position of strength is reached” (ibid.). The exemplary soldier initiates an iterative process in which others follow his example by doing the same thing as he does: taking a stand. In this iterative emulation, an order emerges that becomes stronger and stronger.³¹ This order is neither innately pre-given “in a determinate form” nor inspired by an outside standard: Its strength and specificity *emerge* organically as more and more particulars join in and take their place in it.

The emergent order is aimed at a specific, predetermined ultimate state: Aristotle’s analogy leaves room for ramifications and degrees of freedom. Indeed, the order need not even continually advance to ever more stability and rigidity: individual soldiers could take positions that make the formation as a whole more vulnerable. A rout may occur again—indeed, casting the analogy in the terms of a rout implies that there must have been a former state of stability that the retreating army has lost. The iterative, emergent formation is open to correction, and even breakdown.

This order is also self-referential and self-engendering. Everybody joins in and takes his place with respect to the order that is originating *from that first soldier*: everybody adds to it by repeating his gesture of taking a stand. In taking those stands, each following soldier

³¹ This is by no means guaranteed: the order could also develop into a direction of weakness; and this is a real possibility, as the metaphor speaks of a rout having taken place in a battle in which presumably two parties began in a position of strength and stability.

actualizes what they interpret to have been envisioned, anticipated, and meant by the first soldier and his followers up until then. Through a long chain of implicit reinterpretations, retrievals, and occlusions, the original “intention” of the exemplary soldier keeps exerting itself over the entire order: it is cybernetic. *Induction begins with, and is inconceivable without, a particular that takes the position of an exemplar for other particulars.*

The exemplar thus has an inborn directedness, yet there is no pre-determined unmoving ultimate goal towards which it has to move. In other words, there is no single pre-existing aprioric structure that the exemplar merely activates – the exemplar brings this structure about in a historic process of emergence. The exemplar is not complicit in a deductive order, but thoroughly inductive. It may lead to a deductivism, but it need not. The exemplar is that which initiates, catalyzes, and sustains the process of induction in the Husserlian sense. The exemplar anticipates a universal (more precisely, that there is a universal), takes a stand in the ever-changing flow of experiences, and sets off a process of idealization towards it. This is what Aristotle means when he repeatedly states that the universal is present in the particular, and this is what Husserl means when he speaks about the anticipation inherent in particular experiences.

3. Example and Exemplar

The exemplar has to be guarded against several misunderstandings. First, it is crucial not to mistake it for a metaphor: the exemplar is not a representative that stands in for a universal through a relation of substitution. *The particular does not take the place of the universal* (i.e. the first standing soldier is not replacing the order that he engenders). The exemplar is not a token that has taken the place of something that is “behind,” more prior, or more authentic than the exemplar itself. Conceiving of the exemplar as a metaphor would place it under the

epistemological aegis of a Platonic anamnesis or retrieval, which Aristotle explicitly disavows. There is no need to decipher or unmask the exemplar.

Nor is the exemplar to be understood metonymically, i.e. as hierarchically equal element within a chain of particulars that extends indefinitely in both directions. The first soldier is not (just) one among many. In the same vein, the exemplar is the first member in a structure of particulars, and it engenders that structure. It appears on the scene as an origin and does not derive from other, prior knowledge, which is the other epistemological paradigm that Aristotle rejects in *Posterior Analytics* II.19. To say that the exemplar is neither metonymic nor metaphoric (even though the danger that it is misread in the terms of one or the other always looms large) is to say that it functions according to an altogether different epistemological dynamic.

This dynamic has to be understood through the exemplar's ineluctable status with regard to the cybernetic order it brings about: on the one hand, it is a material member of the formation. It is the initial linchpin of the whole structure, but becomes less and less necessary in its material manifestation as the structure emerges. As such, the exemplar is spatio-temporally located, and an instance among many. Concomitantly, however, it exerts itself over the entire structure across time and space: the exemplar has set in motion an iterative and self-referential process that always, qua iteration, implicitly refers back to it. The exemplar is a particular that institutes a universal. The exemplar is thus the exact opposite of an example: while the example is a specific that simply falls under the aegis of a general that is known to exist before, the exemplar *brings about* the universal that did not exist independently of it before. The exemplar is at once transcendental and ontological. It can only be both at once.

Aristotle closes his short analogy with the pronouncement that the soul indeed has the capacity to undergo this kind of insight from the particular to the universal. "And the soul is

such as to be capable of undergoing this” (100a14). For Aristotle, this transition is not a matter of wishful thinking, but a fact. The human intellect is able to make this jump to the universal. How can Aristotle make this connection? To begin, Aristotle embeds his analogy in a genetic account that moves from perception through an emergence of a structure to a noetic moment of grasping by comprehension: “Thus it is plain that we must get to know the primitives by induction; for this is the way in which perception instills universals” (100b3-5). From there, Aristotle rules out understanding (episteme) as the kind of grasp we have of universals, leaving only comprehension (nous) (100b9-13). Notice, however, that the three stages do not merely succeed each other. Perception is *discriminatory* for Aristotle. This is to say that to perceive means to distinguish something from something, i.e. to see something *as* something: “although you perceive particulars, perception is of universals—e.g. of man, not of Callias the man” (100a17-100b2). The universal is always already distinguished and anticipated in perception. Indeed, without this anticipation, there would be no perception in the first place for Aristotle: it is always already under a noetic pull towards the universal. The moment of grasping by nous, i.e. the noetic gesture is already prefigured in perception itself. Nous is already operative from the moment perception occurs. The assertion that the universal is already present with the particular is a corollary to the more fundamental truth that nous is already operative in perception, and that perception, qua discrimination, is meaningless and insufficiently understood if it is conceived as separable from nous.³²

The empirical and the ideal, the particular and the universal, the ontological and the transcendental are thus not only compatible, but inextricably intertwined for Aristotle. Indeed,

³² This account of induction is far more complex than the simpler conceptions of induction as 1) transition from particular to universal, 2) the inverse of deduction, or 3) mode of scientific discovery. Yet, it is in no way at odds with them: all three versions can be explained and made sense of by this more complex and primal account.

each is inconceivable without the other. Aristotle and Husserl are thus fundamentally aligned in bringing together two realms of thought and knowledge that have been considered severed in the modern era since Descartes. Their parallelism is confirmation for the hypothesis that the lifeworld in which this inductive practice unfolds is neither exclusively an ontological Boden nor a transcendental Leitfaden for science and thinking, but a concept that is supposed to encompass both and to elucidate the transition from the former to the latter. The difference between Aristotle's and Husserl's doctrines is that the former provides an explicit link – nous – where the latter is silent and wrestles with the problem of the attitudinal shift. Aristotle's conceptual fulcrum between the two orders is the exemplar and its attendant inductive process, connecting perception and essential insight, as he indicates in his analogy of the rout. However, if Husserl's and Aristotle's accounts of perception are congruous, and if Aristotle's account shows that such discriminatory perception cannot be conceived other than being suffused by nous, then so must Husserl's! Aristotle's analogy of the rout and its implicit theory of perception, nous, and exemplarity thus promise to hold the key to Husserl's vexing motivational problem.

Why does the first soldier take a stand, despite the overwhelming odds against his survival? He would not be able to give a sufficient reason for his decision, to ground it in prior considerations, truths, and premises. He does not make a rational decision when he takes his stand. He simply takes it. And yet, the reason why he takes a stand is simple: he wants to live! To run away from the overwhelming forces of irrationality, to give himself into the unceasing rush and flow will end in death for certain, *unless* he eventually decides to take a stand. He will only survive for as long as he runs, but this can by definition not be forever. Whether or not his stand is ill-fated does not matter, for before *taking* a chance, there is no chance for him to live at all. In a peculiar inversion of the Pascalian wager, it is earthly life's most basic interest

to establish rationality, an abiding truth within a running stream and constant onslaught of irrationality. To aim for the attitudinal change, to experience inductively, is to be aiming for a deductive order, is to emerge from the realm of doom to the realm of life. Chances at successfully establishing an abiding rational structure may be low, chances of survival in irrationality are definite at *zero*. There is no rational basis for rationality itself, but exactly this baselessness makes the case for rationality all the more forcefully.

1) To live is to induce. 2) To induce is to be motivated towards rationality. 3) To be motivated towards rationality is to be motivated towards the attitudinal shift. Hence, *to live is to be motivated towards the attitudinal shift*. Husserl inscribes this motivation into his inductive notion of life, and indeed makes it inconceivable without it. We cannot pose the motivational question in rationalist terms, i.e. we cannot ask for a rationally motivated origin story of rationality itself. To therefore ask the motivational question, to be puzzled at the fact that such a shift occurred (simply because it cannot be localized or explained in the terms of a pre-given set of reasons) is to misunderstand Husserl's notion of an inherently inductive life. The misunderstanding about Husserl's motivational problem is in the supposition that he delineates an attitudinal shift and then has to provide an explanatory motivation, a question he never explicitly addresses even though he poses it. The reason Husserl does not "address" the problem is because he answers the question in a different way: motivation is inscribed into inductive living towards the deductive from the very start. The entire genesis, self-formalization, and ultimate attitudinal shift occur because the underlying operative motivation is the *sine qua non* for that genesis to occur and unfold in the first place.

To recapitulate: the motivational problem leads us to the Humean problem of induction, and to Aristotle's answer of the exemplar. Rationality can only emerge inductively, i.e. only if consciousness is able to arrest a particular by grasping it as an exemplar. In other

words, consciousness must be disposed towards and trained at recognizing exemplars as origins of inductive chains. This is nothing else than what Husserl attempts to achieve by his (explicitly non-deductive, yet scientific!) method of free variation that seeks to grasp the *eidos*, or essence of a thing it perceives. This is exactly the attitude change that Husserl attempts to bring about in the phenomenological reduction: the central importance of the exemplar in his philosophy is the final piece to complete his elaborate implicit theory of induction.

V. EXEMPLARITY, INDUCTION IN HUSSERL'S THEORY OF LIFE AND INDUCTION: PERFORMING THE EPOCHÉ

Exemplarity is an indispensable component of a doctrine that grasps induction at its root. Without a notion of exemplarity, there could never be an inductive process in the first place. Husserl, as already established, did not explicitly develop his doctrine of induction beyond the sketchy and nascent stages that he reached late in his thinking. However, if Husserl is indeed committed to, and to be localized in, a tradition of thought that places a premium on induction over deduction, and if exemplarity plays such a crucial role in such a thinking, it should be detectable in the late Husserl's phenomenology.

This is indeed the case. In a recent study, Christian Ferencz-Flatz has pointed out that examples provide the crucial intuitive basis for Husserl's notions of phenomenological description and ideation, but that their *modus operandi* is largely unclear (Ferencz-Flatz). Merleau-Ponty, too, in a lecture designed to introduce its audience to the underlying tendencies of Husserl's thought, diagnoses in Husserl's phenomenology a thoroughly inductive thrust that crucially relies on a notion of exemplarity. This notion enriches Husserl's implicit doctrine of induction in comparison to other, more impoverished ones:

Psychology, like physics and the other sciences of nature, uses the method of induction, which starts from facts and then assembles them. But ... that

induction will remain blind if we do not know in some other way, and indeed from the inside of consciousness itself, what this induction is dealing with. (Merleau-Ponty 58)

Induction in the full sense has to anticipate something toward which it induces. It has to encounter particulars as exemplars. To understand how consciousness is disposed towards essential insight, Merleau-Ponty argues that we have to

work out “a sense of examples” to illustrate this insight. What, then, exactly is the relation between this sense of examples and what is called induction? We must here recall the profound remarks of Husserl on induction in general. These were basically opposed to the theory of induction which held sway at the end of the nineteenth century, essentially that of Mill. (68)

Merleau-Ponty discerns this central but unclear role of exemplarity within induction throughout Husserl’s phenomenology.

Merleau-Ponty relies mostly on quotations from Husserl’s 1913 *Ideen I* to argue for this connection. Indeed, Husserl discusses exemplaric consciousness and its training (Einübung) much more explicitly in this earlier work than he does in the *Krisis*. This does not mean that the late Husserl has abandoned his interest in exemplarity. Rather, it has shifted from level of thematic interest to that of an operative paradigm. The supplement to section 9 (a) of the *Krisis* makes a clear connection between inductive thinking and the role of exemplarity:

Exakte Objektivität ist Leistung der Methode, geübt ... als eine Praxis, in der jene unvollkommenen Dingvorstellungen das Material bilden, und zwar in einer gemeinen Denkeinstellung, in welcher von einem exemplarischen Einzelding als Exempel für “irgendein Ding überhaupt” die offen endlose Mannigfaltigkeit seiner immer unvollkommenen aber immer zu vervollkommnenden subjektiven Vorstellungen durchlaufen gedacht wird. (Hua VI, 359)

Husserl hints that he wants to perform an entering into phenomenology with the reader by taking on the “Pfad, den ich selbst gegangen bin” [xxx]. That this performance of method relies on a strategic deployment of exemplars becomes clear from the well-known supplement “The Origin of Geometry,” in which Husserl states outright that his study will proceed in a

way “so daß unsere die Galileische Geometrie betreffenden Probleme und Auslegungen eine exemplarische Bedeutung erhalten” (Hua VI, 365). Indeed, Galileo can only meaningfully be read as an exemplar. Clearly, Husserl is not interested in the historical individual for his own sake, but takes the changes he introduced into scientific practice as a point of departure for a study that quickly moves beyond the specificities and historical facts of that particular individual: Husserl is interested in the attitudinal shift that Galileo brought about with his change in method.³³

Despite temptations to the contrary, Galileo should not be read as a metaphor. While he *can* be read as a representative of an entire new mode of thinking, this, by itself, is neither compelling nor what is at stake for Husserl. Nor should Galileo be read as a metonymy. Husserl acknowledges that Galileo did see himself as an inheritor in a long tradition of knowledge, but makes the emphatic point that Galileo instituted a genuinely new way of seeing the world and of scientifically engaging with it. Galileo set a new beginning, and everybody followed his example (in their own way). His exemplarity This becomes clearer when comparing him to Descartes and Kant, both of whom attempted to become exemplars of philosophy by setting a new beginning, but whom Husserl evaluates as having *failed* in doing so. Both Descartes and Kant wanted to establish a self-transparent, scientific philosophy that was without presuppositions and thoroughly accounted for the transcendental subject (*Krisis* 76, 99), but lost sight of their own fundamental intention. Instead, they succumbed to a blind, self-misunderstanding, deductive rationalism *because* they were under the spell of Galilean thinking (81; see also 83, 93-94). Rather than practicing truly first philosophy by being

³³ The fact that this change in evidentiary structures was never fully conscious to Galileo does not detract from his functioning as an exemplar: he saw himself as inheriting a mathematical tradition from the ancients that he merely amended and improved. The attitudinal shift, i.e. the universal phenomenon (which was itself the transition towards universality) occurred in and through Galileo.

exemplaric thinkers, Descartes and Kant were mere examples of a rationalist worldview that had been instituted by the exemplar of modern science: Galileo. He was able to inaugurate an attitudinal change—Descartes and Kant were not.

Husserl aspires to become an exemplar in the same way: he wants to institute the next attitudinal change, namely from theoretical to phenomenological (*Krisis* 103-04). In other words, Husserl wants to execute consciously and methodically what Galileo achieved unbeknown to himself and what Descartes and Kant fell short of: shift our evidentiary structures and make us see the world differently than we have before. To this end, Husserl introduces the lifeworld, the grounding of thought and practice that Kant, Descartes, and Galileo (as well as Hume) had *all* overlooked and with it, objectivism's ineluctably subjective ground, which we are never supposed to forget.

The lifeworld is mentioned for the first time in section 9 (h) and appears more fully in sections 28, 29, 33, and 34. The lifeworld appears in our view as a particular, i.e. as a concrete entity in material and ontological terms: “[Die Lebenswelt] ist ein Reich eines ganz und gar in sich abgeschlossenen Subjektiven, in seiner Weise seiend, in allem Erfahrbaren, allem Denken, in allem Leben fungierend, also überall unablösbar dabei” (114). The title of section 33 puts it, “ein Teilproblem im allgemeinen Problem der objektiven Wissenschaft” (123). However, again, Husserl extends the scope of this problem by arguing that it arises out of subjective performances of inductive experience that inhere in subjectivity as such (130). By the end of his exposition of the problem of a science of the lifeworld, Husserl has reversed his stance: the penultimate subsection of section 34 is entitled, “Das Problem der Lebenswelt anstatt als Teilproblem vielmehr als philosophisches Universalproblem” (135). This reversal is not an oversight. Rather, Husserl has taken the reader on a path in order to point to the task that philosophy as phenomenology has to take on: the transcendental epoché. The lifeworld points

beyond its own particularity to a problem of universal, all-encompassing implications: if we perform the transcendental epoché, everything will necessarily begin to look different for us.

In other words, the lifeworld is an exemplaric concept for Husserl. It is never for the sake of understanding the lifeworld as such, (which would merely repeat the mistake of the objectivist thinking he opposes). Rather, it serves to rehearse a much more fundamental move, to which Husserl devotes the remainder of part III.A: understanding and executing the epoché, i.e. of abstention from preconceived notions and prejudices. From this perspective, it is perfectly understandable that Husserl drops the unfinished investigation of the lifeworld and transitions to an analytic of the transcendental epoché in section 35 and following: the study of the lifeworld, which encapsulates all of our preconceived notions, is an exemplaric study of how we are stuck in such preconceived notions, and how a thematization of the unthematized constitutes a first step beyond those preconceptions.

The following sections document Husserl's increasing concern to properly grasp the transcendental epoché and its execution. Husserl emphasizes that executing it results in a radical turn away from the natural, naïve attitude in which one once lived (151). How exactly this is to be done raises huge problems for him (151-53): "Hier bieten sich ... verführende Irrwege, d.i. Weisen, die Durchführung der Epoché zu verstehen, welche sicher nicht zum Ziele führen" (151-52). „[W]ie groß hier die Versuchung zu Selbstmissverständnissen ist und wieviel, ja schließlich das wirkliche Gelingen einer Transzendentalphilosophie, an der selbstbesinnlichen Klarheit bis ins Letzte hängt, werden die weiteren Überlegungen zeigen“ (156). Husserl struggles here to maintain the proper perspective. Going down specific, particular paths of inquiry has to be explicitly postponed to keep the study as a whole on track towards what it intends: a change in attitude. Husserl is thus forced to only indicate those paths (see 163-68). Indeed, Husserl himself alerts the reader in section 49 of the

“exemplarische Enge der ausgeführten Analysen” (170). Getting bogged down in all the details and ramifications that his thinking imply would catch him in a myopia that would take the “object” of study as exactly such an *object* and lose the perspective that he aims to achieve but has only dimly in view so far. From an exemplaric beginning twice executed, Husserl’s thinking begins to feel the strong pull of anticipation—only towards what is not fully clear.

This forward-rushing thinking does not mean, however, that it merely needs to hasten to its intended insight, for everything to fall into place. Husserl’s thinking is not immune to correction. On the contrary, towards the end of part III.A, in section 52, Husserl recognizes that his science of the lifeworld runs the risk of repeating the naïve mistake of an objectivist understanding of it (178-79), an abstention from interestedness as a blind repetition of a solipsistic turn away from the world (179-80), and the problem of how exactly transcendental subjectivity is to be understood (181-82). Husserl thus ends part III.A with a section titled “Die prinzipielle Korrektur unseres ersten Ansatzes der Epoché durch Reduktion derselben auf das absolut letztlich fungierende Ego” (190) before he begins his path into phenomenology anew.

The second time Husserl attempts to enter into phenomenological thinking is from the clarified vantage point of transcendental subjectivity and its temptation to fall into objectivism (203). Husserl’s thinking zig-zags: having rushed forward, it stops, evaluates what it has achieved so far, and identifies the need for a new, more informed, and clearer beginning. It becomes clear that Husserl was *also* not interested in the transcendental epoché as such, i.e. the abstention from all kinds of particular intentionalities towards specific intentional objects (233). Such particular intentionalities can easily be misconstrued in objectivist terms. Husserl *again* performs the move of seeing those intentionalities exemplarily, trying to grasp to which

each of them points: he seeks to find out what constitutes abstention from intentionality as such, i.e. he is interested in grasping what constitutes an epoché *as such*.

This can only occur once a few exemplaric epochés of a “lower” level have been performed. As a result, the epoché has to be practiced again and again in order to ultimately see in what such a gesture as such consists. Husserl compares the psychological-phenomenological reduction to all previous ones in the following way:

Die phänomenologische Psychologie erschließt sich ihrem Sinne nach in verschiedenen Stufen, weil die phänomenologische Reduktion selbst – und das liegt in ihrem Wesen – ihren Sinn, ihre inneren notwendigen Forderungen, ihre Tragweite nur in Stufen erschließen konnte. Jede erforderte neue Reflexionen, neue Besinnungen, die ihrerseits nur möglich waren durch das Selbstverständnis und die geübte Leistung der anderen Stufen. (250)

A little further down, Husserl concludes that the kind of engagement with the world for which he aims cannot simply be achieved in an attempt at an epoché that is wholly without preparation: this kind of perceiving

[war] vor der Einführung der eigentümlichen Methode der phänomenologischen Epoché überhaupt nicht möglich. Demnach musste der phänomenologisch sich Einstellende erst sehen lernen, Übung gewinnen und in der Übung zunächst eine rohe und schwankende, dann immer bestimmtere Begrifflichkeit von seinem und anderen Eigenwesentlichkeiten erwerben. (251)

Phenomenology comes into its own inductively. Again, a concept (this time, epoché) appears on the scene as a specific particular (its transcendental version), but it clarifies itself as the universal problem. The science of the lifeworld leads to the exemplaric epoché (the transcendental one) in order to set off an iterative, inductive practice that seeks to advance to the noetic moment in which the essence of each of those particular instances of abstention is grasped. This may happen, or it may not, but such a moment is anticipated, from the moment the lifeworld comes into view (and perhaps even before that). This leads to the “vollbewusst

geübte Methode, die ich ... die *phänomenologisch-psychologische Reduktion* nenne" (239), a method that Husserl already anticipates in the Galileo section (46).

Husserl's method is one of training himself and his readers in always seeing more in a particular than what is immediately perceived. His cryptic indication that "Die 'gesehenen' Dinge sind immer schon mehr als wir von ihnen 'wirklich und eigentlich' sehen. Sehen, Wahrnehmen ist wesensmäßig ein Selbsthaben ineins mit Vor-haben, Vor-meinen. Alle Praxis mit ihren Vorhaben impliziert Induktionen" (51), applies not merely to perception, but is prototypical of human thinking and life as such. Husserl utilizes this inherent inductivity of living thinking to show us in action what he means. What we apprehend is not merely what we apprehend, nor is it a metaphor or a metonymy. It points beyond itself exemplarily.

In revealing Galileo's exemplarity, Descartes' and Kant's failures to become exemplars (i.e. new beginnings of different attitudes and evidentiary structures), and proposing his own philosophy as a correction to their approaches, it is clear what Husserl is attempting to achieve: he wants to become an exemplar of scientific thinking as *only* Galileo managed to become. Descartes and Kant's failures show how such a task is *not* achieved: in simply repeating the gesture of instituting a deductive, rationalist worldview without a full conscious awareness and understanding of that fact. This blindly repeats Galileo's gesture and therefore does not lead to a new beginning: it simply does what Galileo has done already. Instead, Husserl can become an exemplary thinker by grasping and reiterating what was truly exemplary by Galileo: the fact that he inaugurated an attitudinal shift (not the particular, specific attitudinal shift that he factually inaugurated). This conscious grasp of Galileo lets Husserl problematize the issue of an attitudinal shift as such, bring it into view, and study it.

In other words, Husserl is able to retrieve a meaning from Galileo that nobody before Husserl had been able to retrieve: what is at stake is not the shift to a deductive worldview per

se, but the shifting of attitudes as such. Husserl can therefore himself become an exemplar, i.e. institute a new beginning in philosophy for others to follow, by making the past appear in a different light. Husserl's repetition of the Galilean act is executed in a way that fundamentally changes our understanding of him. We see Galileo (and Descartes, Hume, and Kant) differently after Husserl. Galileo, who instituted an ineluctably different way of seeing the world was an exemplar in *his* time. It is now Husserl who does that by aspiring to be an exemplar in his own historical moment. Husserl suggests nothing short of the audacious proposition that Galileo himself (not just Kant, Hume, or Descartes) bore the seeds of phenomenology. Indeed, one could go so far as saying that Husserl argues that experience itself, with its inductive tendencies and inclination to discriminatorily and exemplarily see things as more than what they are, to be inclined to phenomenology. This addresses the problematic question of an exemplar's: it is never set completely *ex nihilo*, as there is always some past to which it reacts. Nonetheless, the exemplar sets an origin in that it makes the past reveal for the first time what it "truly has been all along." This is how new beginnings and an abiding intention, transmitted through tradition can be reconciled in Husserl's thought.

Perception—science—philosophy as phenomenology: the *Krisis of the European Sciences* is the last attempt at an all-encompassing philosophy, yet, not in the trappings of a system. Rather, it tries to master its sprawling subject matter in a history of attitudinal changes, i.e. of seeing things in different lights in different eras. Husserl's inductive philosophy of exemplarity thus does indeed make headway toward a new kind of universal reason: one that sees coherence and continuity in the contingencies of history, yet avoiding the pitfalls of a barren deductivism. Husserl's inductive philosophy is an always emergent train of thought of attitudinal shifts, epochés, abstentions from being, etc., constantly questioning and clarifying itself and seeking to set new beginnings. Husserl took a stand against all the philosophies of

his time that made no pretensions of questioning their own underlying assumptions and views, be it logical empiricism, Neo-Kantianism, psychologism, historicism, or irrationalist life philosophy.

To live is to induce, and induction never occurs without the iterative repetition of a noetic moment in which a universal is anticipated by grasping the exemplarity of a given particular. The attitudinal shift is thus always already motivated and prefigured in life's most basic operation. While being predisposed in this way, consciousness need not ever factually undergo such a shift: its occurrence is entirely historically contingent. While its noetic minutiae are not entirely clear, what *is* clear is that it has occurred at least once in history. If systematic and mathematical thinking are thus tied back to an inductive practice based on such a noetic moment of exemplary insight, the sciences cannot do without the humanities: this moment cannot be ignored or banned from scientific thinking lest it fundamentally misunderstand itself. It is incumbent upon the humanities to serve as a reminder of this fact—the lingering obscurities of exemplarity and its origins are not reasons to debunk it, but call for a continued reckoning and clarification. In humanistic and scientific thinking alike, it is imperative to identify and contemplate its originary, exemplary moment that set off its self-perpetuating inductivity. This is to continue Husserl's unfinished and unfinishable work. His infinite task is not one that can ever be considered completed. Husserl's last work remained unfinished, and this is probably the most telling indication.

WORKS CITED

Primary Texts

- Aristotle. [APo] *Posterior Analytics*. Second ed., trans., commentary J. Barnes. Oxford: Clarendon Press, 1993. Print.
- Aristotle. *Metaphysics*. In: McKeon, Richard (ed.). *The Basic Works of Aristotle*, New York: Modern Library, 2001. 689-935. Print.
- Carnap, Rudolf. [Carnap a] "Statistical and Inductive Probability." In: Brody, Baruch A. (ed.). *Readings in the Philosophy of Science*. Englewood Cliffs, N.J.: Prentice Hall, 1970. Print. 440-450. Print.
- — —. [Carnap b] "On Inductive Logic." In: *Readings in the Philosophy of Science*. 451-477. Print.
- Hume, David. [Treatise] *A Treatise of Human Nature*. Ed. L. A. Selby-Bigge, second ed., rev. P. H. Nidditch. Oxford: Clarendon Press, 1979. Print.
- — —. [Enquiries] *Enquiries Concerning Human Understanding and Concerning the Principles of Morals*. Introd. L. A. Selby-Bigge, third ed., rev. P. H. Nidditch. Oxford: Clarendon Press, 1975. Print.
- Husserl, Edmund. *Logische Untersuchungen Bd. 1. Prolegomena zur reinen Logik*. 2. umgearbeitete Auflage. Halle (Saale): Max Niemeyer Verlag, 1913. Print.
- — —. *Formale und transzendente Logik. Versuch einer Kritik der logischen Vernunft*. Halle (Saale): Max Niemeyer Verlag, 1929. Print.
- — —. [E&U] *Erfahrung und Urteil: Untersuchungen zur Genealogie der Logik*. Prague: Academia Verlagsbuchhandlung, 1939. Print.
- — —. [Hua VI] *Husserliana VI: Die Krisis der europäischen Wissenschaften und die transzendente Phänomenologie*. Hrsg. Walter Biemel. 2. Auflage. Haag: Nijhoff, 1962. Print.
- — —. *Die Krisis der europäischen Wissenschaften und die transzendente Phänomenologie. Eine Einleitung in die phänomenologische Philosophie*. Hrsg. Elisabeth Ströker. 3. Auflage. Hamburg: Meiner Verlag, 1996. Philosophische Bibliothek Band 292. Print.
- — —. "Beilage XIII. Induktion in der Welterfahrung und die Konstitution der orientierten Erfahrungswelt als Welt mit Erde und Himmel." In: *Husserliana XXXIX: Die Lebenswelt. Auslegungen der vorgegebenen Welt und ihrer Konstitution. Texte aus dem Nachlass (1916-1937)*. Hrsg. Rochus Sowa. Dordrecht: Springer, 2008. 184-189. Print.
- — —. [Krisis] *Die Krisis der europäischen Wissenschaften und die transzendente Phänomenologie. Eine Einleitung in die phänomenologische Philosophie*. Hrsg. Elisabeth Ströker. Philosophische Bibliothek Band 641. Hamburg: Meiner Verlag, 2012. Print.

Secondary Texts

- Barnes, Jonathan. "Introduction." In: Aristotle. *Posterior Analytics*. Second ed., trans., commentary J. Barnes. Oxford: Clarendon Press, 1993. Print.
- Bayer, Greg. "Coming to Know Principles in 'Posterior Analytics' II 19." *Apeiron: A Journal for Ancient Philosophy of Science* 30.2 (1997): 109-142. Print.
- Blumenberg, Hans. "Das Lebensweltmißverständnis." In: Blumenberg, Hans. *Lebenszeit und Weltzeit*. Frankfurt a.M.: Suhrkamp, 1986. 9-68. Print.
- — —. "Theorie der Lebenswelt." In: Blumenberg, Hans. *Theorie der Lebenswelt*. Hrsg. Manfred Sommer. Berlin: Suhrkamp, 2010. 7-108. Print.
- — —. "Die Lebenswelt als Thema der Phänomenologie." In: *Theorie der Lebenswelt*. 109-132. Print.
- — —. "Lebenswelt und Wirklichkeitsbegriff." In: *Theorie der Lebenswelt*. 157-180. Print.
- — —. "Lebenswelt und Technisierung unter Aspekten der Phänomenologie." In: *Theorie der Lebenswelt*. 181-224. Print.
- Brody, Baruch A. *The Confirmation of Scientific Hypotheses. Introduction*. In: *Readings in the Philosophy of Science*. 376-382. Print.
- Burnyeat, M. F. "Aristotle on Understanding Knowledge." In: Berti, Enrico (ed.). *Aristotle on Science: The Posterior Analytics*. Proceedings of the Eighth Symposium Aristotelicum, Padua, 1978. Print.
- Bussey, W. H. "The Origin of Mathematical Induction." *The American Mathematical Monthly* 24.5 (1917): 199-207. Print.
- Cajori, Florian. "Origin of the Name 'Mathematical Induction'." *The American Mathematical Monthly* 25.5 (1918): 197-201. Print.
- Carr, David. "Husserl's problematic concept of the life-world." In: Moran, Dermot, Lester E. Embree (eds.). *Phenomenology. Critical Concepts in Philosophy. Volume I: Phenomenology: Central Tendencies and Concepts*. London: Routledge, 2004. 359-374. Print.
- Claesges, Ulrich. "Zweideutigkeiten in Husserls Lebensweltbegriff." Claesges, Ulrich, Klaus Held (Hrsg.). *Perspektiven transzendentalphilosophischer Forschung. Für Ludwig Landgrebe zum 70. Geburtstag von seinen Kölner Schülern*. Den Haag: Nijhoff, 1972. 85-101. Print.
- Derrida, Jacques. *Edmund Husserl's Origin of Geometry: An Introduction*. trans. John P. Leavey, Jr. New York: Harvester Press, 1978. Print.
- Engberg-Pedersen, T. "More on Aristotelian Epagoge." *Phronesis* 24.3 (1979): 301-319. Print.

- Ferencz-Flatz, Christian, "Das Beispiel bei Husserl." In: *Tijdschrift voor Filosofie* 73 (2011): 261-286. Print.
- Fink, Eugen. "Operative Begriffe in Husserls Phänomenologie." In: *Zeitschrift für philosophische Forschung* 11.3 (1957): 321-337. Print.
- Føllesdal, Dagfinn. "The *Lebenswelt* in Husserl." In: Hyder, David and Rheinberger, Hans-Jörg (eds.). *Science and the Life-World: Essays on Husserl's 'Crisis of the European Sciences'*. Stanford: Stanford U Press, 2010. 27-45. Print.
- Freudenthal, Hans. "Zur Geschichte der vollständigen Induktion." In: *Archives internationales d'histoire des sciences* 6 (1953): 17-37. Print.
- Friedman, Michael. "Science, History, and Transcendental Subjectivity in Husserl's *Crisis*." In: *Science and the Life-World*. 100-115. Print.
- Groarke, Louis. *An Aristotelian Account of Induction: Creating Something from Nothing*. Montreal: McGill-Queen's U Press, 2009. Print.
- Hamlyn, D. W. "Aristotelian Epagoge." *Phronesis* 21.2 (1976): 167-184. Print.
- Heß, Walter. "Erfahrung und Intuition bei Aristoteles." *Phronesis* 15.1 (1970): 48-82. Print.
- Lee, Nam-In. "The Pluralistic Concept of the Life-World and the Various Fields of Phenomenology of the Life-World in Husserl." *Husserl Studies* 36 (2020): 47-68. Print.
- Levin, Michael. "Induction and Husserl's Theory of Eidetic Variation." *Philosophy and Phenomenological Research* 29.1 (1968): 1-15. Print.
- Mall, Ram Adhar. "Der Induktionsbegriff. Hume und Husserl." *Zeitschrift für philosophische Forschung* 29.1 (1975): 34-62. Print.
- Merleau-Ponty, Maurice. "Phenomenology and the Sciences of Man." trans. John Wild. In: Merleau-Ponty, Maurice. *The Primacy of Perception, and Other Essays on Phenomenological Psychology, the Philosophy of Art, History, and Politics*. Ed. James M. Edie. Evanston, IL: Northwestern U Press, 1964. 43-95. Print.
- Moran, Dermot. *Husserl's Crisis of the European Sciences and Transcendental Phenomenology. An Introduction*. Cambridge, UK: Cambridge U Press, 2012. Print.
- Murphy, Richard T. "Husserl's Relation to Hume." *Research in Phenomenology* 9 (1979). 198-223. Print.
- — —. "Husserl's Relation to British Empiricism." *The Southwestern Journal of Philosophy* 11.3 (1980). 89-106. Print.
- Nagel, Ernest. "Carnap's Theory of Induction." in: *Readings in the Philosophy of Science*. 478-496. Print.

- Orth, Ernst Wolfgang. “‘Lebenswelt’ als eine unvermeidliche Illusion? Husserl Lebensweltbegriff und seine kulturpolitischen Weiterungen.” In: Preyer, Gerhard, Georg Peter, Alexander Ulfig. (Hrsg.) *Protozoziologie im Kontext. ‘Lebenswelt’ und ‘System’ in Philosophie und Soziologie*. Würzburg: Königshausen & Neumann, 1996. 28-40. Print.
- Palermo, James. “Apodictic Truth: Husserl’s Eidetic Reduction Versus Induction.” *Notre Dame Journal of Formal Logic* 19.1 (1978). 69-80. Print.
- Sepp, Hans Rainer. “Transzendente Theorie und Praktik.” In: Sepp, Hans Rainer. *Praxis und Theoria. Husserls transzendentalphänomenologische Rekonstruktion des Lebens*. Freiburg (Breisgau), München: Alber, 1997. 245-279. Print.
- Stegmüller, Wolfgang. “Das Problem der Induktion: Humes Herausforderung und neue Antworten.” Lenk, Hans (Hrsg.) *Neue Aspekte der Wissenschaftstheorie*. Braunschweig: F. Vieweg, 1971. 13-74. Print.
- Thompson, Wayne N. *Aristotle’s Deduction and Induction: Introductory Analysis and Synthesis*. Amsterdam: Rodopi, 1975. Print.
- Unguru, Sabetai. “Greek Mathematics and Mathematical Induction.” In: *Physis* XXVIII (1991): 273-289. Print.
- Upton, Thomas V. “A Note on Aristotelian Epagōgē.” *Phronesis* 26.2 (1981): 172-176. Print.
- Vacca, G. “Maurolycus, the First Discoverer of the Principle of Mathematical Induction.” *Bulletin of the American Mathematical Society* 16.2 (1909): 70-73. Print.
- Vickers, John. “The Problem of Induction” In: *The Stanford Encyclopedia of Philosophy*. Summer 2010 Edition. Web.
<http://stanford.library.usyd.edu.au/archives/sum2010/entries/induction-problem/>
 (last accessed February 5, 2021).
- Watson, John. “Aristotle’s Posterior Analytics: II. Induction.” *The Philosophical Review* 13.2 (1904): 143-158. Print.

CHAPTER 3: FORMALIZATION'S STRATEGIC PARADOXES IN KAFKA'S DISCOURSE OF GUILT³⁴

I. THE VEXATIOUS QUESTION OF GENRE

“Faßt Du Dein Urteil über mich zusammen, so ergibt sich, daß Du mir zwar etwas geradezu Unanständiges oder Böses nicht vorwirfst ..., aber Kälte, Fremdheit, Undankbarkeit. Und zwar wirfst Du es mir so vor, als wäre es meine Schuld” (144.15-21, underlining in the original). Franz Kafka begins the letter to his father with a cautious, moderate act of rebellion. He says “no” to his father’s “unaufhörliche Vorwürfe,” and “yes” to a shared basic human innocence. Kafka admits that he is cold, estranged, even unfair to his father. But what Kafka refuses to accept is the implicit ascription of *guilt* that comes with each of those Vorwürfe. He will not let his father pronounce him *prima facie* guilty for his actions. By the same token, Kafka expressly abstains from ascribing guilt to his father. Kafka claims that he is only concerned with defending his own basic innocence, not with questioning anyone else’s. Importantly, in pursuit of his rebellious but moderate quest for innocence, Kafka relies on logic. His letter attempts to be a rigorous, methodical *proof* of his innocence.

The peculiar form, aim, and logic of the *Letter to the Father* have left scholars puzzled over the significance and place within Kafka’s larger oeuvre. The *Letter* is often called a key text to understanding Kafka, a “Schwellentext, der die beiden Räume Lebenszeugnisse und Literarisches Werk gleichzeitig trennt und doch miteinander verbindet” (Unseld 189-90). Yet, it is precisely its opaque position between biography and literary creation that has perplexed

³⁴ This chapter was originally published—although substantially different in some parts—as Kuchne, Tobias. “Paradox and Kafka’s ‘Lawyerly Ruse’: A Systems Theoretical Study of *The Letter to His Father*.” *Journal of the Kafka Society of America* 37/38 (2013-14). 89-100. Print. The journal and its website seemed to be defunct at the time of submission. It was therefore not possible to identify the proper recipient for a copyright permission request.

critics and readers to this day. From Max Brod, who insisted that the *Letter* was a biographical piece that was intended to reach the father, but published it with Kafka's literary writings in the 1953 collection *Hochzeitsvorbereitungen auf dem Lande*, to Daniel Weidner, who, in the most recent Kafka handbook, identifies the *Letter's* "faktual-fiktionalen Zwitterstatus" as a source of scholarly vexation (294), critics have repeatedly agonized over determining the *Letter's* genre and pertinent method of analysis.

Despite the undisputed importance of the *Letter*, there have thus been few studies that make it their central focus (see Weidner 294, Brune 75, Stölzl 520). Instead, the *Letter* is often referenced within larger Kafka studies and classified under various headings. Classifications range from biographical (Wagenbach, Gibian, Sokel) to psychoanalytical (Rattner, Schaefer), sociohistorical (Müller-Seidel, Anz, Stölzl), literary (Brune), metaphysical (Politzer 1965), existential (Honegger), and deconstructive (Deleuze and Guattari). Other critics intermix various genre ascriptions without problematizing their relationship (Pfeiffer, Politzer 1953). Each such attempt either seeks to reduce the *Letter* to the status of a non-fictional, biographical document or a fictional, literary text, or declares the two strands irreconcilable (Hasselblatt 155). Carlo Brune succinctly states in his study on the *Letter*: "Es hat den Anschein, als würde die Ambivalenz von biographischen und literarischen Ingredienzen im Brief nahezu allen Analysen ... letztlich zum Verhängnis" (93). Stölzl similarly summarizes:

Entsprechend der inselhaften Selbstgenügsamkeit der verschiedenen Schulen der Kafka-Deutung gibt es zwischen den einzelnen Interpretationsmodellen eigentlich keinen Dialog und keinen aus fortschreitender Diskussion aufgebauten gesicherten Forschungsstand. (520)

This chapter proposes an alternative route. Instead of treating it as a representative of a particular genre, it proposes to study the *Letter's* formal operations. This chapter, in other words, proposes to treat the *Letter* as a piece of communication and a logical proof. The cross-currents of those two operations make for the *Letter's* peculiar effect, which this chapter will

develop in two steps. First, it will show that Kafka's proof runs awry, not because it is not rigorous enough, but because it is *too* rigorous. This proves the opposite of what he claims he wants to prove, namely that he is inescapably mired in a basic condition of guilt. His self-justification is turning into a Vorwurf. Second, the rigorous but flawed logic is designed to evoke a response from his father, i.e. to initiate communication with him. Specifically, Kafka hopes to elicit a self-justification from his father that mirrors those that Kafka provided in the *Letter*. In eliciting this response, Kafka turns the deadly combination of the rigor and failure of his logic on his father and draws him into that same mire of guilt.

To read the *Letter* without guidance by a genre-specific theory, it is necessary to orient oneself in the text. The introductory sentences provide a first clue on how to proceed:

Liebster Vater, / Du hast mich letzthin einmal gefragt, warum ich behaupte, ich hätte Furcht vor Dir. Ich wußte Dir, wie gewöhnlich, nichts zu antworten, zum Teil eben aus der Furcht, die ich vor Dir habe, zum Teil deshalb, weil zur Begründung dieser Furcht zu viele Einzelheiten gehören, als daß ich sie im Reden halbwegs zusammenhalten könnte. (143.1-8)

Franz is afraid of his father, and he claims to rely on the written form of the letter in his attempt to explain this fear. However, a read-through of the *Letter* reveals that Franz is indeed capable of both recalling a vast amount of "Einzelheiten" and stringing them together into a historically and thematically structured narrative. There is no attempt to explain any fear either: indeed, fear is saliently absent throughout the *Letter*. Instead, Franz seems to state the aim he pursues with the *Letter* much more accurately a few paragraphs later. Franz reconstructs his father's view of their relationship as estranged and cold. While Franz readily admits to his lack of affection, he decidedly rejects that he is, in any way, guilty for this coldness, as his father seems to imply. Franz now presents the *Letter* as a proof of his innocence:

Diese Deine übliche Darstellung halte nur soweit für richtig, daß auch ich glaube, Du seist gänzlich schuldlos an unserer Entfremdung. Aber ebenso gänzlich schuldlos bin auch ich. Könnte ich Dich dazu bringen, daß Du das anerkennst, dann wäre – nicht etwa ein neues Leben möglich, dazu sind wir

beide viel zu alt, aber doch eine Art Friede, kein Aufhören, aber doch ein Mildern Deiner unaufhörlichen Vorwürfe. (144.25-145.5)

The proof begins immediately after this passage, which makes it a more apposite indicator of the *Letter's* communicative context: occasioned by a recent remark by the father, it is an attempt to alleviate the persistent condition of the father's reproaches to reduce the deep friction between them. The inconsistency between the two cited passages highlights the difficulty that the reader faces in situating the *Letter*. These qualms can be allayed somewhat by looking at a letter that Franz drafted before *The Letter to the Father* and that provides more insight into its genesis:

Es war nicht das erste Mal, daß ich einen Vorwurf von Dir Vater, hörte und Du hast mir ihn nicht nur durch Türen, sondern auch schon unmittelbar ins Gesicht gesagt und es war auch nicht nur der Vorwurf der Teilnahmslosigkeit allein ... trotzdem wurde ich gerade an jenem Abend, von dem Vorwurf, den ich nicht einmal ganz deutlich gehört hatte, trauriger noch, als sonst. Ich suchte lange einen Ausweg, bis mir endlich im Bett einfiel, ich könnte Dir ... einen Brief schreiben, und alles darin erklären. In der Freude, die mir dieser Einfall machte, fielen mir gleich 100 Dinge ein, die ich zu schreiben hätte, und *überdies noch ein System, das sie vollkommen überzeugend machen sollte*. (Unsel 207-08, emphasis mine)

This passage makes it clear that the *Letter* does not aim to explain Franz's fearfulness, but to combat his father's constant reproaches: a particular reproach struck Franz more deeply than usual and spurred him to write a letter. Moreover, the idea did not present itself in a disjointed manner, nor did it overtax his mental capacities: Franz could "immediately think of 100 things" he would present, along with "a system to make them convincing." These two claims are absent from the introductory lines of the actual *Letter*, where they have been replaced by their opposites. One can thus read the opening passage of the *Letter* as a concealment of its aim to carefully—and methodically—disprove the father's frequent reproaches. Against its historical context of the father's constant insinuations of guilt, the *Letter* has to be understood as a rigorous, comprehensive, and methodical proof of Franz's

innocence. Franz begins the *Letter* with a list of propositional statements that aim for exhaustiveness (cf. Kafka 143.17-144.24). Indeed, in letters to Milena, Kafka exhorts her: “verstehe beim Lesen alle advokatorischen Kniffe, es ist ein Advokatenbrief,” and: “Der Brief ist doch zu sehr auf sein Ziel hin konstruiert” (Unseld 192). But as Brune critically remarks: “Auf semantische Feinheiten und die ausgeklügelten ‘advokatorischen Kniffe’ des Briefes wird zwar stetig verwiesen, ohne diesem jedoch Taten folgen zu lassen” (Brune 75). The following will be an attempt at such a delivery. The first step is thus to take the *Letter* at its word and follow its logic. It is only then that we can study what it does.

II. APPLYING GÖDEL’S CALCULUS TO KAFKA’S LOGIC

1. Justifying the Use of Formal Language

Kurt Gödel devised a toolkit that enabled him to explore the boundaries of logic. Gödel’s crucial move was his idea of gödel numbering, a calculus that enabled the creation and attribution of one set of logical expressions to another set of logical expressions. The method of gödel numbering takes a set of expressions—the *object language*—and assigns exactly one number to each expression. The assignment procedure is unambiguously defined by criteria that I lay out below. The resulting set of gödel numbers is the *meta-language* that “talks about” expressions in the object language, which is to say, it “interprets” them.³⁵ Every statement in the object language corresponds to exactly one gödel number in the meta-language *and vice versa*. Each member in a statement-gödel number pair can be seen as talking about the other: object statement and gödel number are surrogates for each other. This approach has the advantage of sidestepping problems of self-referentiality. Since object

³⁵ For more information on what follows, refer to George S. Boolos, Richard C. Jeffrey. *Computability and Logic*. Third Edition. Cambridge University Press, Cambridge 1989, p.170.

statements and meta-statements are equivalent, object statements that “talk about themselves” can be analyzed by referring to their gödel number.

I propose to use gödel numbering to formalize and follow the logic that Kafka attempts to use against his father. Kafka interprets his father’s statements in a way that meets all the structural criteria of assigning gödel numbers to an object language. The only difference is that Kafka does not assign numbers to propositions, but propositions to propositions. The way in which Kafka goes about this project, however, is equivalent to the method of gödel numbering, which gives us two languages that stand in exactly the same relationship as does a set of gödel numbers to its object language. From the interplay between gödel numbers and their object language, Kurt Gödel went on to derive the *Diagonal Lemma*. The Diagonal Lemma states that, for any sentence G in the object language, there exists some other statement B (in the object language) that makes a statement about the gödel number of G . According to the Diagonal Lemma, G and B can only be simultaneously true or simultaneously false. In other words, the Diagonal Lemma shows that a set of expressions is not independent from its set of interpretive expressions. (This will be spelled out in more detail below.) In the case of the meta-language that Kafka creates for his father’s set of propositions, I will rely on the Diagonal Lemma to show that 1) Kafka’s proof of his innocence is not only contradictory, but that 2) it is the proof itself that *engenders* and betrays his guilt.

I base my formalization of Kafka’s logic on the opening four paragraphs of the “Letter.” This may seem to grant unwarranted attention to a disproportionately short passage, but the passage is not chosen at random. The introduction and Kafka’s reference to it in the final paragraph frame the body of the letter. The introduction claims to make the entire rest of letter a proof of Kafka’s innocence. The framing passage sets up the parameters for this proof. Formalizing the framing passage is to take Kafka at his word about what the letter as a

whole attempts to achieve. With his letter, Kafka stages a wider scheme that comes to fruition with the deliberate collapse of a logic that needs to be established but has to be tucked away and effaced in a short passage. The implications of this collapse can only be sufficiently cashed out if the logic itself is fully explicated first.

There is an intuitive qualm about an attempt to narrow a narrative text to a formal language such as propositional logic. Texts are not propositions, but (*at least*) a web of interconnected propositions. However, several facts justify my move. Kafka's father – or rather, the image of the father that Kafka paints in his letter, which is all we are concerned with for our purposes (Kafka: “ich rede ja nur von der Erscheinung, in der Du auf das Kind wirktest,” 148.11-12) – views the world in terms of simple propositions with immutable truth values. “Dir hat sich die Sache immer sehr einfach dargestellt” (143.14), Kafka states as he speaks of his father's “geistige Oberherrschaft. Du hattest Dich allein durch eigene Kraft so hoch hinaufgearbeitet, infolgedessen hattest Du unbeschränktes Vertrauen zu Deiner Meinung. ... Deine Meinung war richtig, jede andere war verrückt, überspannt, meschugge, nicht normal” (151.25-152.6). If the father's thought consists of statements with a clear truth value, the field of his language should be expressible in logical propositions.

I will simplify and limit my formalization of Kafka's language by formalizing only the basic moves that he makes: he reinterprets every claim the father makes about his son's actions to carry an implicit ascription of guilt; and he counters such accusations by arguing that, “action x by Kafka shows that Kafka is not guilty.” (This will be the statement $B(x) = x \Rightarrow \neg g$, which will be of particular interest when we apply the Diagonal Lemma.) Formalizing in greater detail than that would serve no additional purpose, while bogging reader and writer down in unnecessary detail.

2. Formalizing the Father's and Kafka's Language

Kafka concludes his opening paragraph thus: “Faßt Du Dein Urteil über mich zusammen, so ergibt sich, daß Du mir zwar etwas geradezu Unanständiges oder Böses nicht vorwirfst ..., aber Kälte, Fremdheit, Undankbarkeit. Und zwar wirfst Du es mir so vor, als wäre es meine Schuld” (144.15-21). According to Franz, the father operates with a clandestine double register. To the father, every communication by Franz eo ipso betrays his lack of love, which he repeatedly stresses. Everything the father says to Kafka truly says two things: the propositional content of the statement, and the added statement, “You are guilty.”

My formalized language consists of few components. I distinguish between two sets of propositions: the object language EX of the father and the meta-language KX by Kafka. EX is a set of propositions in which each proposition makes a claim on some action x done by Kafka. KX is the set of propositions that interpret the propositions in EX . Note that we do not have direct access to EX , i.e. the father’s statements, which forces us to exercise care in formalizing them. Note in particular that we do *not* know whether or not the father actually ascribes guilt to his son: it is *Kafka* who claims that the father does. All we know is that the father’s expressions are a family of propositions, each about some action x by Kafka, and most likely the same in their basic structure. Call those expressions $S(x)$, where x is the only free variable.³⁶ It need not concern us too much *what* the father says about his son’s actions. What matters is that he claims *something* about what Kafka does, and once x is replaced by a particular action c , $S(c)$ becomes a *sentence*, i.e. it is either true or false.

The body of the letter features quoted sentences of the father (although mostly approximated in their quotation) that seemingly diverge from the requirement that they refer

³⁶ A free variable is a placeholder for a concrete filler that can take its place. When there are no free variables, the expression (or formula) becomes a *sentence* and acquires a *truth value*. In our case, the formula $S(x)$ covers the whole set of sentences the father can utter: just insert some specific action for x .

to some action x by Kafka. I exclude those sentences from the father's language that refer to some other subject ("Er [der lungenkranke Kommissar] soll krepieren, der kranke Hund," 173.5; "Die Gottselige [Irma] hat mir viel Schweinerei hinterlassen," 182.18-19), since they do not bear on Kafka's proof of his innocence vis-à-vis the father. Remarks such as "Immer alles contra" (159.21) and the father's long denigration of Kafka's fiancée (207-08) do fit the model. Sarcastic comments such as "Kauf dir was dafür!" to the young Kafka (153.12) and the request to put the new book on the nightstand (192.14) do not. I reiterate that I am basing my formalization on the first four paragraphs of the letter, i.e. the formalization that Kafka suggests when he says, "Dir hat sich die Sache immer *sehr einfach dargestellt*," (143.14, emphasis mine) before he goes into his wholesale summary of the statements his father routinely flings at him (ibid.). Moreover, it is no stretch to suggest that, for Kafka, all the quotations referred to above ultimately do refer to something he did, and that they do mean to imply his guilt.

KX is the set of *interpretations* of the statements in EX . Kafka gives a blanket interpretation to his father's statements: they all *not only* make a statement about action x , but they *also* claim that Kafka is guilty. Call those interpretations of the father's expressions ' $S(x)$ ' $= S(x) \wedge g$, i.e. for any of his (Kafka's) actions x , Kafka claims, the father does something more than saying something about x : the father makes a claim about Kafka doing some action x , *and* the father says that Kafka is guilty. The assignment process has to meet three criteria (see Boolos & Jeffrey 170): (1) Each gödel number is assigned to only one expression in EX . (2) Every expression in EX has exactly one gödel number, and that gödel number can be calculated in a finite number of steps. (3) For any number, there is a decision procedure of finitely many steps that tells us *whether* that number belongs to some expression in EX , and if so, *which* expression the number belongs to. These three conditions amount to saying that i) there are two sets of expressions, ii) each member of each set is paired up with exactly one

member from the other set, and iii) there is an unambiguous way to compute the partner of each expression in finitely many steps.

There is nothing that prevents us from having a set of propositions be interpreted in this way by another set of propositions, as long as the three conditions above are still met. Kafka's interpretations KX of the statements in EX pass this test. For each particular action x , Kafka assigns the statement " $S(x) \wedge g$ " to the father's statement $S(x)$. Every proposition in EX thus corresponds to exactly one proposition in KX , since each proposition in the object language uniquely determines its interpretation by reappearing *in* it. Each proposition in KX that corresponds to some proposition in EX can be obtained in finitely many steps: add " $\wedge g$ " to $S(x)$ for any particular x , to obtain ' $S(x)$ '. This fulfills condition (2). It is also clear that if a proposition in KX interprets some proposition in EX , it cannot interpret any other proposition in EX , which fulfills condition (1). To check whether a proposition in KX has a corresponding proposition in EX , see if it ends in " $\wedge g$," in which case the corresponding proposition in EX is the same as the one chosen expression from KX , minus the phrase " $\wedge g$ ". By that process, we ascertain both *whether* an expression in KX has a corresponding expression in EX , and also *what* that expression is. This fulfills condition (3). Kafka performs a linguistic move that is equivalent to gödel numbering. As mentioned above, the upshot of this formalization is that an object statement can "talk about itself" without running into the pitfalls of self-referentiality: Statements in EX that refer to themselves can be treated as referring to their equivalents in KX . Indeed, one of the reasons why Gödel devised gödel numbering was to find a way to avoid the paradoxes that result from self-referential statements such as, "This sentence is false."

3. Conditions for the Diagonal Lemma to be Derivable

I intend to apply Gödel's Diagonal Lemma to EX and KX . In non-formulaic terms, the lemma states this: Take two statements from the object language EX , and call them G and $B(x)$. In addition, take $\ulcorner G \urcorner$, which is the gödel interpretation of G . $B(x)$ has a free variable, x , which can be filled with either a statement from EX , or a gödel number. If we fill $B(x)$ with $\ulcorner G \urcorner$ to obtain $B(\ulcorner G \urcorner)$, B is now saying something about the gödel interpretation of G . We now want to prove that, for *every* $B(x)$, there exists a G , such that *G is true if and only if $B(\ulcorner G \urcorner)$ is true*. This is the Diagonal Lemma.

To prove the Diagonal Lemma for two sets of propositions EX and KX , Gödel showed that two criteria (in addition to the three discussed above) had to be fulfilled: (1) the propositions can be *diagonalized*, and that (2) the propositions can represent a specific function called *diag*, which will be defined below.

In all generality, a diagonalized expression is: $\exists x((x = \ulcorner A \urcorner) \wedge A)$. (“There exists an x such that x is equal to $\ulcorner A \urcorner$; $\ulcorner A \urcorner$ and A are true.”) This says that two conditions obtain: there exists some statement $\ulcorner A \urcorner$ (the Gödel interpretation of A), and both $\ulcorner A \urcorner$ and A hold true at the same time (Boolos and Jeffrey 172). This becomes interesting when A is an expression that has only one free variable, namely x . In this case, we can write A as $A(x)$, and since x becomes $\ulcorner A \urcorner$ for a diagonalized statement, we obtain $A(\ulcorner A \urcorner)$. Diagonalized statements are precisely what Gödel wanted to construct to escape the pitfalls of self-referential statements such as “This statement is false.” Instead, self-referential statements are recast as diagonalized statements that talk about their own interpretations: “The gödel statement associated with this statement is false” is now a statement that can have its *own* truth value without contradicting itself.

For the propositions in our set EX , diagonalization is both possible and meaningful. Look at the statements $S(x)$ first. They have exactly one free variable, namely x . In the

diagonalization of $S(x)$, $\ulcorner S(x) \urcorner$ therefore takes the place of x , so that, for any particular action x , the diagonalization of $S(x)$ becomes $S(\ulcorner S(x) \urcorner) = S(S(x) \wedge g)$. We stipulated in the beginning that all we know about x is that it is an action by Kafka. It is perfectly possible to make x the interpretation of a statement by the father (in fact, this is the only action we *see* Kafka perform through the letter). It is within the reach of the father's language as circumscribed by Kafka to make a statement $S(x)$ about the particular action $x = \ulcorner S(x) \urcorner$. It is conceivable that the father can make a statement about Kafka saying that the father makes statements and implies guilt with them.³⁷ Imagine the father having read the letter (which *is* the container of interpretations of the father's statements as described above) and making a statement about Kafka having written it. Kafka himself acknowledges this possibility at the end of the letter when he offers a hypothetical response by his father.

Diagonalization has drawn our attention to an important subset of propositions within the object language *EX*. A straightforward question to ask then is *if diagonalized propositions, too, can be interpreted*: does Gödel's calculus allow us to assign a gödel number to diagonalized statements, and is Kafka's interpretation of his father's statements able to perform an equivalent move?

The answer to both questions is yes. A diagonalized statement is still a statement in the object language, and the gödel numbering calculus is required to give a unique interpretation for *every* object statement. Within Gödel's calculus, the assignment process of a gödel number to a diagonalized statement is performed by a function we call *diag* (see Boolos and Jeffrey 172). Two sets of propositions are able to represent *diag*, if, for every diagonalized expression in the object language, there exists a Gödel interpretation of that expression in the

³⁷ Note that $S(\ulcorner S(x) \urcorner)$ is again a *family* of statements, and it is the family of *all* diagonalizations. Each of them has one free variable and becomes a sentence for a particular action x .

meta-language. In our case, this means that for every $S(\ulcorner S(x) \urcorner)$ in EX , there must be a $\ulcorner S(\ulcorner S(x) \urcorner) \urcorner$ in KX . This is surely the case. Since Kafka sees an indictment in *every* statement by the father about an action x by Kafka, he would have no qualms to say $\ulcorner S(\ulcorner S(x) \urcorner) \urcorner$, i.e. that his father implies guilt with his statement $S(\ulcorner S(x) \urcorner)$. (“I [the father] claim something about the fact that you [Kafka] interpret my statements about your actions x to imply guilt.”) EX and KX are able to represent the function *diag*.³⁸

Without laying out Gödel’s construction of *diag* in detail, what interests us is that he showed that it is always possible to create a mechanism that pairs statements G from the object language with statements $B(\ulcorner G \urcorner)$ from the object language such that G and $B(\ulcorner G \urcorner)$ must both be true or both false. In other words, gödel numbering allows to tie the truth value of statement G to the truth value of statement B about G ’s gödel interpretation.

4. Applying the Diagonal Lemma to Kafka

The question now becomes which statement $B(x)$ interests us. We can pick any $B(x)$ we want, since the Diagonal Lemma applies to *any* $B(x)$ in the object language. Consider again the goal Kafka claims to pursue with his letter: he wants to prove his innocence. Kafka wants to prove the claim that, for some action x that he picks, x implies that he is not guilty. Hence, $B(x) = x \Rightarrow \neg g$, i.e. “If Kafka did action x , then he is not guilty.” This is the statement that Kafka wants to press into the object language. He wants to make the father acknowledge the truth of this claim, i.e. he wants to make it one of the statements the father could and would utter. Now, according to the Diagonal Lemma, $G \Leftrightarrow B(\ulcorner G \urcorner)$. G must be a sentence from the object language EX , while $\ulcorner G \urcorner$ must be its interpretation from the meta-language KX . Hence,

³⁸ We need not be concerned with the construction of the actual function. All we need is the possibility of its existence, i.e. the function’s representation via its results.

for a specific action ϵ , there must be some sentence $G = S(\epsilon)$ for which the Diagonal Lemma holds: $G \Leftrightarrow B(\neg G)$ becomes $S(\epsilon) \Leftrightarrow B(\neg S(\epsilon))$, i.e. $S(\epsilon) \Leftrightarrow ((S(\epsilon) \wedge g) \Rightarrow \neg g)$. The father makes some claim about some action ϵ by Kafka. This claim is true if and only if it is also true that Kafka's act of interpreting this claim by the father to carry with it an ascription of guilt implies that Kafka is not guilty. The two are either simultaneously true or simultaneously false.

The Diagonal Lemma demands that the statement $S(\epsilon) \Leftrightarrow ((S(\epsilon) \wedge g) \Rightarrow \neg g)$ be true. In other words, the expression on each side of the double arrow must both be true, or both be false. We therefore need to assess which truth values are possible for its elements $S(\epsilon)$ and g . Suppose, to begin with, that $S(\epsilon)$ is false, i.e. that the father makes an *incorrect* statement about action ϵ by Kafka. Then $(S(\epsilon) \wedge g) \Rightarrow \neg g$ as a whole is false as well. This is because, if $S(\epsilon)$ is false, then the antecedent $(S(\epsilon) \wedge g)$ of the conditional is false regardless of the truth value of g , since “false and something” make for a false. However, *any* conditional that has a false antecedent is true. $(S(\epsilon) \wedge g) \Rightarrow \neg g$ is true regardless of the truth value of g . We thus know: Supposing that $S(\epsilon)$ is false makes $(S(\epsilon) \wedge g) \Rightarrow \neg g$ true, regardless of whether Kafka is guilty or not. The biconditional $S(\epsilon) \Leftrightarrow ((S(\epsilon) \wedge g) \Rightarrow \neg g)$ as a whole thus says: “false iff true”—a plain logical paradox.

This establishes the first conclusion: There must be at least one statement $S(\epsilon)$ by the father that *cannot be false*. If what the father says is *always* incorrect, Kafka's proof of his innocence violates the Diagonal Lemma. We must rule out the possibility that, for any ϵ , $S(\epsilon)$ is false: There exists some ϵ for which $S(\epsilon)$ is true.

Hence suppose that $S(\epsilon)$ is true and examine $(S(\epsilon) \wedge g) \Rightarrow \neg g$. If we assume that g is true, then $\neg g$ is false. The antecedent $(S(\epsilon) \wedge g)$ of the conditional is then true, but the conclusion false, which makes the conditional as a whole false, and again the biconditional

winds up stating a contradiction: “true iff false.” To avoid this paradox, we must rule out the possibility that g is true, which is to say: we must rule out that Kafka is guilty. Therefore, suppose that g is false and $\neg g$ is true. In that case, the antecedent of the conditional becomes false, the conditional as a whole becomes true, and the biconditional $S(\epsilon) \Leftrightarrow ((S(\epsilon) \wedge g) \Rightarrow \neg g)$ says “true iff true.”

Using Kafka’s language, the structure of his claim about his innocence, and the Diagonal Lemma, the cumulative result becomes this: for Kafka’s proof of his innocence to be possible, there must be 1) at least one statement $S(\epsilon)$ by the father that is true, and 2) $\neg g$ must be true, i.e. Kafka must actually be innocent. The Diagonal Lemma rules out all other possibilities.

5. Interpreting the Result of the Diagonal Lemma

If the Diagonal Lemma suggests that the only possibility is that Kafka *is* innocent, Kafka’s proof seems to have been successful. The formalization suggests exactly what Kafka claims to have set out to do in the beginning of the letter: the father makes certain statements about his son’s behavior to him, and while Kafka accepts the truth of those statements, he wants to prove that he incurs no guilt for the actions the father talks about.

However, this convenient (and not particularly exciting) result only holds up as long as we remain inside the Gödelian calculus. Once we explore the implications of the fact that $S(\epsilon)$ must be true for some particular action ϵ , paradox returns on a level beyond the logical one and implodes Kafka’s proof. From our above definition of the family of statements $S(x)$, we do not know what the father is specifically claiming about Kafka’s actions. We only know that the father makes propositional statements that are all equal to each other, except for the particular action substituted for the variable x . Two mutually exclusive but exhaustive cases

therefore apply to both $S(x)$ in general, and to $S(c)$ in particular: the father either pronounces his son guilty with *all* of his statements, or he does *not* pronounce his son guilty with *any* of his statements.

Suppose the father does pronounce the son guilty with all of his statements $S(x)$, in which case $S(c)$ pronounces him guilty. Implication 1) of the Diagonal Lemma was that $S(c)$ must be true, which implies that the father *correctly* declares his son guilty. But implication 2) of the Diagonal Lemma was that Kafka is *not* guilty. $S(c) \Leftrightarrow ((S(c) \wedge g) \Rightarrow \neg g)$ then yields the paradox that Kafka can prove his innocence if and only if he is guilty.

It must therefore be the case that the father does *not* pronounce his son guilty with his statements. This avoids the paradox of the biconditional, but at the cost of making Kafka's proof a vacuous endeavor: he can only prove his innocence *if the father never meant to say that his son is guilty*. Kafka's proof can only be coherent if the father never brought up the question of guilt. Kafka then proves something that never required proof in the first place.

It is then *Kafka*, with his interpretation of his father statements as " $S(x) \wedge g$ " that introduces the notion of guilt.³⁹ This is not an outright paradox, but it is a stunning result. How can Kafka convincingly defend himself against a claim of guilt that *he* has created? The performance of proving his innocence introduces the guilt he is arguing against. Within the logic of the proof, this is impossible to see, but it emerges as soon as we engage the proof from an observatory meta-level (this time in a hierarchical sense). A major implication of Gödel's First Incompleteness Theorem is that every logical system has a point of necessary

³⁹ Indeed, we can trace this transition in the very first paragraphs, in which Kafka summarizes his father's propositions and then seamlessly moves into the register of "Urteil," Schuld," and "Vorwürfe" (1202/3).

incompletion. Gödel's calculus provides a method to uncover those blind spots,⁴⁰ and in the case of Kafka's letter, the blind spot turned out to be the guilt lurking behind the assertion of innocence.

III. NACH LOGIK DIE SINTFLUT: THE LOGIC OF VORWURF, RESPONSE, AND GUILT

1. Eliciting a Response

It would be naïve, however, to think that Kafka was not aware of the guilt-engendering performativity of his proof. In the first few pages, Kafka characterizes his father's statements as Vorwürfe. A Vorwurf is a statement that conveys a rebuke of the addressee. The rebuke itself tends to be left unsaid, but is implied in the statement's formulation, expression, and the context in which it is brought up. "You do not love me, you are estranged from me," is, on the face of it, a mere statement of fact. Kafka himself seems to have been aware of the distinction when he comments: "Das ist natürlich kein Vorwurf, nur Feststellung einer Tatsache" (158.15-16). The addressee must extrapolate the implicit rebuke—otherwise the statement never becomes a Vorwurf to the addressee, regardless of the addresser's intention. A Vorwurf only begins to function as a Vorwurf once the addressee takes up the gauntlet and acknowledges that there *is* a rebuke. The father never seems to have done that, which led to his "rätselhafte Unschuld und Unangreifbarkeit" (161.4-5). If the addressee does not take a Vorwurf as a Vorwurf, she does not feel guilty. It turns out that the addresser's intention hardly matters in this context. Whether the gauntlet is fake or real, a Vorwurf is a relational phenomenon: a statement becomes a Vorwurf to the addressee in the addressee's *interpretive*

⁴⁰ The blind spot of the performativity of the system uncovering the blind spot of the initial logical system of course remains. Blind spots of object system and meta-system need to be kept distinct.

act. This is the case in Kafka's letter to his father. The interpretive act, the gödel numbering of the object statement, matters far more than the object statement itself (which is why our limited knowledge about the content of the father's statements was sufficient for our purposes).

Regardless of the position that the interpreter takes to the rebuke, she can only acknowledge it as such if she takes it to refer to something that *is* rebuked, real or imagined. To interpret something as a rebuke and react to it is already an acknowledgement that there *is* something to be reacted to, that there is a battle to be fought, a defense to be leveled. To interpret something as a Vorwurf presupposes an awareness of guilt as a concept and the recognition that the question of guilt applies to the specific situation. Calling something a Vorwurf betrays this awareness. Guilt can never be *fully* argued away, as the argument constitutes an irreversible act of acknowledgement. The only pure state of innocence is one of ignorance of guilt (a "rätselhafte" Unschuld indeed to someone like Kafka who only knows a life of guilt), a state that is made inaccessible by the very attempt to prove its existence.

Guilt thus arises not in isolation, but through discourse and interpretation. Guilt is intrinsically distributed among two (or more) persons. There must be a subject uttering a statement, and there must be another subject interpreting and recognizing it as a Vorwurf. Kafka creates this type of discourse in his letter: his doubling of language according to the Gödelian calculus creates the possibility for such a discourse between two parties to take place.⁴¹ It requires an "outside" source of a statement, so that there is an uncertainty and a need to interpret for the subject that cannot penetrate to the inside of that outside source. It is the mark of an innocent mind that the statements of an outside or an other do not spark its

⁴¹ To be sure, the two subjects need not be two different persons. All that is required is that there are two different *subjectivities* that engage in this interaction: Kafka may well have made up the figure of the father.

desire bridge the gap between the two: the innocent mind does not *perceive* the problem of such a gap, or the need for interpretation. Guilt can never be explicitly stated and then be taken as such: it is the *making* explicit that creates it. An ignorantly innocent person could even be the addressee of the statement: “You are guilty for doing x ,” and the innocent person could respond “Yes I am” without *feeling* guilty: the ignorantly innocent’s guilt is just another brute fact for him that, as such, does not create the feeling of guilt.

The claim that guilt can only *arise* in discourse and through interpretation is somewhat imprecise and only partially correct, since it incorrectly suggests that it is possible to trace a temporal origin of guilt. Once a subject recognizes guilt, whether on account of a particular instance or in general, this guilt is recognized as something already present. It is recognized as something that has already been there before its recognition. There can be a point in time at which guilt is *discovered*, but its discovery as something that has already been there makes guilt apply atemporally. The “origin” of guilt is elusive and pre-historical.

Franz Kafka tries (or feigns to try) to pin down the absence guilt in the propositional statements of a logical calculus. What the calculus does not capture is the performance of its own operations. While it avoids the pitfalls of self-referential statements by creating a parallel interpretive language, the act of interpretation establishes a relationship between the two that is its own nest of paradoxes. Simple languages encounter a singularity in self-referential statements. Doubled languages of the type that Kafka constructs presuppose the existence of the concept about which the interpretive act makes inferences. The doubled language encounters *its* singularity if the attempted inference is the non-existence of that concept.

2. Kafka’s Lawyerly Ruses

Consider the full implication of Kafka's move: the discourse of guilt draws in everyone who takes a position to a Vorwurf. Anyone who can be tricked into taking a position is therefore duped out of the realm of innocence and is irretrievably introduced into the world of guilt. Kafka attempts to do precisely that to his father, as a couple of crucial moves in the latter part of the letter reveal: (1) Kafka not only professes to prove that he is innocent, but he claims the same about his father as well. His proof, however, shows that he cannot be innocent. (2) Kafka bases his proof on interpretations of statements that the father makes about Kafka's actions. Throughout the letter, Kafka also makes statements about *his father's* actions. Given the symmetrical relationship that Kafka attempts to stage between himself and his father, one cannot but help to think that Kafka hoped to induce his father to *interpret* his son's statements as Vorwürfe as well. While a collection of factual and propositional statements on the surface, Kafka's letter is structured in a way to seduce the father into taking a position to them by interpreting his son's letter *as* a Vorwurf, and to draw the father into the discourse of guilt.

"[D]as Gelingen des ganzen Briefes" (199.7) is thus staked on one question: can Kafka get his father to engage in the discourse of guilt? Can he make the father respond and say: "but I am *not* guilty!"? The main body of the letter, beginning with the fifth paragraph, pursues this aim. Contrary to a first impression, the body of the letter is intimately wound up with Kafka's logical calculus. The author does not spend his effort on a tangential matter, but on the all-important thing: kicking the calculus into action by making his father speak the words that will push him beyond the point of no return. These are the "advokatorischen Kniffe" that Kafka asked Milena to find in his "Advokatenbrief" (Unseld 192).

There is no reason for Kafka not to overdo it. According to Kafka's account, the father seems quite resistant to receiving thinly veiled messages, such as when he gave him Franklin's youth memoirs to read:

[D]as ist ja selbstverständlich und wieder gar kein Vorwurf, sondern nur eine Abwehr Deiner Vorwürfe. Du hast letzthin Franklins Jugenderinnerungen gelesen. Ich habe sie Dir wirklich absichtlich zum Lesen gegeben, aber nicht, wie Du ironisch bemerktest, wegen einer kleinen Stelle über Vegetarianismus, sondern *wegen des Verhältnisses zwischen dem Verfasser und seinem Vater*, wie es dort beschrieben ist. (190.10-18, emphasis mine)

Even if the father seems to have an unconscious inkling of having to acknowledge something, he does not make the step, as Kafka indicates in the beginning (143), and when he thematizes his father's turning away from Judaism in response to Kafka's rising fervor as a subconscious acknowledgment of his lacking devotion (190-91). To make his father take the irreversible step, Kafka fires a salvo of provocations at him that fall into at least four different categories.

The first method in which Kafka attempts to evoke a defensive reaction from his father is by his tone of voice. The sheer crassness and vividness of some of his anecdotal evidence must, upon any commonsensical reading imply a Vorwurf of guilt. A few examples:

Dich aber hörte und sah ich im Geschäft schreien, schimpfen und wüten, wie es meiner damaligen Meinung nach in der ganzen Welt nicht wieder vorkam. Und nicht nur schimpfen, auch sonstige Tyrannei. Wie Du zum Beispiel Waren, die Du mit anderen nicht verwechselt haben wolltest, mit einem Ruck vom Pult hinunterwarfst – nur die Besinnungslosigkeit Deines Zorns entschuldigte Dich ein wenig – und der Kommiss sie aufheben mußte. (172.22-173.3)

Unverständlich war immer Deine vollständige Empfindungslosigkeit dafür, was für Leid und Schande Du mit Deinen Worten und Urteilen mir zufügen konntest, es war, als hättest Du keine Ahnung von Deiner Macht. (154.22-26)

Tiefer gedemütigt hast Du mich mit Worten wohl kaum und deutlicher mir Deine Verachtung nie gezeigt. (206.1-2)

Regardless of what the context is—Kafka can insist on his father's innocence as much as he likes—these statements are Vorwürfe, and they aim at evoking feelings of guilt with the father.

If the father is as unresponsive to implicit suggestions, the second plank of Kafka's agenda to incite a defense from the father may be more promising. Throughout the letter, Kafka stages his relationship with his father by reiterating the following points: the father is strong, powerful, and naïvely self-assured, while the child is weak, defenseless, and insecure. The son is entirely dependent (Kafka explicitly adduces his failed marriage as the most recent and most resounding example of his inability to establish his independence; 206-08). Everything that the son does is therefore in a close relationship to the father, even an expression or extension of the father himself: "Ich stand ja in allem meinem Denken unter Deinem schweren Druck. . . Alle diese von Dir scheinbar unabhängigen Gedanken waren von Anfang an belastet mit Deinem absprechenden Urteil" (152.23-24; 152.26-153.1).⁴² The father, however, is inconsistent, which makes him an overpowering giver of standards that are by nature incomprehensible and impossible to meet. Kafka speaks of his world that is infinitely far away from the father's; the former receives his commands from the latter and can *never* fully satisfy them (156-57). Due to their unique relationship, Kafka had to fail in everything he attempted, had to feel guilty about everything, but was powerless to change anything, since his father was the *source* of all of his actions. In summary: for everything Kafka does, the father is the origin, and guilt is the result. The inference that the father is the source *of* guilt (and thus guilty) is the obvious last step in the syllogism that Kafka never concludes.

3. Dropping the Pretense

Or *almost* never. If the father is neither amenable to a suggestive tone nor to completing a syllogism, Kafka is even willing to breach his declaration that the father is guiltless. Consider

⁴² Notice the suggestiveness of the particle "ja" here: Kafka is indicating that, whatever he did, it could not have been otherwise.

this phrase from the longer quote above: “nur die Besinnungslosigkeit Deines Zorns *entschuldigte* Dich ein *wenig*.” His frenzy excuses (“de-guils,” “un-guils”) the father only *partially*. The process of excusal is incomplete, i.e. there must be “some guilt left.” During Kafka’s anecdote of the missed sex education talk, he remarks: “An sich war es ja unverständlich, daß Du mich so verurteiltest, nur alte Schuld und tiefste Verachtung Deinerseits konnten mir das erklären” (204.11-14). If we posit that the father’s condemnation *can* be comprehended and explained, then one of the causes must be guilt. For Kafka to openly assert his father’s guilt, we would need to rely on an assumption that we cannot ascertain, but suffice it to say that Kafka’s language is more than merely suggestive here. Even such a strong attribute of Kafka’s language would be charitable with this statement:

Dagegen sollst Du zwar auch in diesem Punkt, ebenso wie ich, an Deine Schuldlosigkeit glauben, diese Schuldlosigkeit aber durch Dein Wesen und durch die Zeitverhältnisse erklären, nicht aber bloß durch die äußeren Umstände. (189.23-27)

Kafka points to the staging of their relationship in the letter without calling it by its name. He makes the father’s guiltlessness a pretense of consensus between the two. He is only willing to grant this stage device, however, if the father can, at least partially, derive his guiltlessness from his own constitution, i.e. if he can show guiltlessness not to be in total contradiction with his self. Kafka all but challenges his father to *justify* himself, to make the gesture on which he stakes his whole letter. However, Kafka does not stop at blatant suggestions. He openly calls his father guilty once when he says: “ich mußte ja, wie ich schon sagte, das an ihnen gutmachen, was Du unter meiner Mitverantwortung im Geschäft an ihnen verschuldet hattest” (184.14-18). This statement does not require external assumptions or extensive interpretation: Kafka has called his father guilty.

As soon as this acknowledgement happens – as soon as he can make the father pick up the gauntlet – Kafka has leveled the playing field. This is the goal that Kafka has pursued

all his life. His letter is his last hope: his attempt at marriage has recently failed, and with it, his last shot to achieve “die schärfste Selbstbefreiung und Unabhängigkeit. ... Ich wäre Dir ebenbürtig, alle alte und ewig neue Schande und Tyrannei wäre bloß noch Geschichte” (209.4-9). Through writing, Kafka pursues what was barred to him through other means. What is at stake in his letter is thus nothing short of realizing his last opportunity to attain the most fundamental objective of his life. The last step that separates him from achieving it is to make his father assert his innocence. From this perspective, the length of the letter is hardly surprising. Rather, it is surprising that Kafka did not write an entire book of examples that attempt to evoke such a reaction from his father.

Whether or not the father can *convince* himself that he is innocent (if Kafka’s portrayal of him is remotely accurate, it seems likely that he does), he cannot return to being fully, ignorantly innocent. The letter, while a transparent, fair, and impartial proof at first blush, does not turn out to be a proof after all, but a linguistic mudwrestling match in order to push he father beyond a point of no return, out of the comfort zone from which he (innocently?) wielded his destructive power.⁴³ It is thus no longer surprising that the letter is preoccupied with making this step happen. The logical machinery, on the other hand, that is then set into motion need only be indicated – and would need to efface itself – to work effectively.

However, a careless reader, as Kafka makes his father out to be, may still overlook suggestive tone of voice, nearly finished syllogisms, and even blatant accusations. To make absolutely sure to push his father over the edge, Kafka fires his most deadly shot at the end,

⁴³ Interestingly, if we are correct about the father’s ignorant innocence that the letter seems to cast him in, Kafka makes a true statement about his father’s innocence. Yet, the act of reading the statement by the father would turn it into a false one, as the statement would raise his awareness to an all-encompassing and atemporal guilt.

where he provides a fictitious reply by the father. It seems as if Kafka thwarts his effort by unmasking it in his father's fictitious voice:

[Du willst] auch mich von jeder Schuld freisprechen. Natürlich gleingt Dir das letztere nur scheinbar (mehr willst du ja nicht), und es ergibt sich zwischen den Zeilen trotz aller 'Redensarten' von Wesen und Natur und Gegensatz und Hilflosigkeit, dass eigentlich ich der Angreifer gewesen bin. (214.19-24)

Kafka topples his explicit project of proving his innocence through the voice he gives to his father. His father's self-defense morphs into offense at this point. Instead of justifying himself, the father uncovers his son's stratagem:

Jetzt hättest Du also schon durch Deine Unaufrichtigkeit genug erreicht, denn Du hast dreierlei bewiesen, erstens daß Du unschuldig bist, zweitens daß ich schuldig bin und drittens daß Du aus lauter Großartigkeit bereit bist, nicht nur mir zu verzeihn, sondern, was mehr und weniger ist, auch noch zu beweisen und es selbst glauben zu wollen, daß ich, allerdings entgegen der Wahrheit, auch unschuldig bin. (214.25-215.5)

The father stages a counterattack by exposing the ruse of the *Letter's* double-dealing formalism and condemning it. He reinstates his own superiority within the register of guilt by privileging his "more honest" mode of struggling and by uncovering Franz's underlying motivation:

Das könnte Dir jetzt schon genügen, aber es genügt Dir noch nicht. Du hast es Dir nämlich in den Kopf gesetzt, ganz und gar von mir leben zu wollen. Ich gebe zu, daß wir miteinander kämpfen, aber es gibt zweierlei Kampf. Den ritterlichen Kampf, wo sich die Kräfte selbständiger Gegner messen, jeder bleibt für sich, verliert für sich, siegt für sich. Und den Kampf des Ungeziefers, welches nicht nur sticht, sondern gleich auch zu seiner Lebenserhaltung das Blut saugt. ... Wenn ich nicht sehr irre, schmarotzerst Du an mir auch noch mit diesem Brief als solchem. (215.5-14; 216.20-22)

Why does Franz let the father lash out so violently, calling his son "Ungeziefer" and a "Schmarotzer," someone who has always fed off of him and is continuing to do so with this letter? Franz overshoots the mark by letting the father expose the *Letter's* entire mechanism, its hidden designs, and the effects it achieves. At the last moment, the father reassumes his overpowering position with a few deft and devastating rhetorical strokes. Franz is unable to

rebut his father (how could he deny such a disarmingly accurate exposure?) and contents himself with a weak *ad hominem*: “Darauf anworte ich, daß zunächst dieser ganze Einwurf, der sich zum Teil auch gegen Dich kehren lässt, nicht von Dir stammt, sondern eben von mir” (216.23-25).

What matters is that he puts words of self-defense *into* his father’s mouth. The father does not even have to make the step of asserting his innocence – Kafka does it *for* him. All that the father has to do now is *agree* with this fictitious quote of his, and he has irretrievably entered the discourse of guilt. It is precisely because this fictitious quote claims something ringing so true that makes it so dangerous. From this angle, it becomes clear why Kafka bothers little to respond to this apparent attack on himself: his actual motive of leveling the playing field of guilt has been achieved, which will truly make “Leben und *Sterben* leichter” (217.10-11, emphasis mine) for both: Kafka cannot overcome the guilt that pervades every fiber of his existence and forever weds him to the overawing influence of his father. He can, however, reflect the infection back to his father by, in the manner of Ungeziefer, constructing a self-defeating logic and grafting it into his father’s thinking.

IV. COMING TO REST?

Has Kafka won the fight? To say this would be a premature conclusion. A pan-destructive logic, in solidarity with neither the addressee nor the addresser, cannot leave a victor and a loser once the dust has cleared. Kafka does not hope for the possibility of a “neues Leben,” only for “eine Art Friede, kein Aufhören, aber doch ein Mildern Deiner unaufhörlichen Vorwürfe” (145.3-5). When he makes his concluding remarks, Kafka reiterates this gist of his claim by saying that his fictitious quotation has created something close enough

to the truth, “daß es uns beide ein wenig beruhigen und Leben und Sterben leichter machen kann” (217.10-11).

Kafka hopes for an alleviation of the viciousness that has marked their relationship. The annihilation of one party by the other, the knightly dueling between two independent parties that the father supposedly engages in, is no longer possible. Kafka’s letter has forged an inextricable link between the two—not through his repeated insistence on his dependence on the father, but by tricking the father into the discourse of guilt in which both parties presuppose each other in everything they say. As such, the fight of the vermin, of one parasitically feeding on the other, is the only possible fight left. Kafka has carried the day in irreversibly imposing his rules of fighting, but has concomitantly debunked the hope of the fight ever ending in a (knightly) victory for him or his father.

The infinite deferral of the struggle’s outcome seems to thwart the possibility for the Beruhigung that he speaks of in the last sentence. Take note, however, that the struggle was vicious precisely because there was a misconceived hope that one of them could win out over the other. Despite their unequal strength, Kafka oppressed the father just as much the father oppressed Kafka. Passages such as “ehe Dich Deine Kinder, besonders ich, enttäuschten und zu Hause bedrückten” (147.9-11), “[wir waren] einander so gefährlich” (147.17), “So littest Du, so litten wir” (164.3), “Erst später begriff ich, daß Du wirklich durch die Kinder sehr littest” (164.20-21), and “man dachte nur an den Kampf, den Du mit uns, den wir mit Dir führten” (176.17-18) are frequent in the letter, and they suggest a fight that was vicious because it was – at least partially – waged on the father’s terms. As long as there was a conceived possibility for an ultimate victor, fighting was brutal and unceasing, constantly reaffirming “[d]ie Unmöglichkeit des ruhigen Verkehrs” (159.5).

In Kafka's terms of fighting, physical strength is no longer a factor. He has thus truly achieved the leveling of the playing fields that he could not attain through marriage. As both feed upon each other, neither can live without the other. Realizing this does not lead to a ceasing of the fight – they *live* off of each other – but it diffuses some of its viciousness. The fight becomes a calm gesture in the resignation that they cannot interact any differently. Beruhigung sets in, and living and dying becomes a little easier, as they replenish themselves by draining the other.

Taking Kafka at face value has led to a calculus that unmasked itself and Kafka's letter as performing the opposite of what the author claimed. Instead of proving his innocence, he proves both his father's and his. Kafka wins by making nobody win, and after his task is completed, he concludes his letter with calming words that require again to be taken at face value. Interpretations that champion one contender over the other, painting either the father as the domineering tyrant or Kafka as the insolent pesterer could also make peace with each other and find Beruhigung, if they started out from taking Kafka at his word and letting him lead where he takes them.

To argue about one's guilt is to no longer be innocent. Kafka's logic is immensely powerful, not because it is a set of straightforward syllogisms, but because it moves, with all possible rigor, to the point where it unravels in a paradox that we cannot get out of once we have fallen into it. He dismantles the domain of basic human innocence so thoroughly that one cannot help but think that it is a mere ruse when Kafka claims it is the conclusion he argues for.

Kafka certainly does not view basic human innocence as an unassailable value, as he seems to imply in the beginning of the letter. Kafka also certainly exercises no moderation in his reasoning. It is not his aim to defend a human boundary line that cannot be crossed. Rather,

he ruthlessly pushes against and crosses them, others' as much as his own. It is his aim to invade and eradicate the domain of innocence in every subject, be they strong or weak, be they already guilty or still innocent, be they someone else or himself. The only means to evade the unwinnable struggle is not to read his letter. Kafka's mother knew this and returned the letter to her son. The Beruhigung of the letter never occurred because Kafka's safe haven in the family thwarted him.

The Letter relies on a lawyerly ruse to unfold a paradox via a formalism that impels the *Letter* forward to the point of the formalism's rupture and the paradox's explicit reappearance. It seems to me that this is Kafka's performance of the "Offenlegung des hinter allen literarischen Metaphern und Chiffren ... wirkenden *Prinzip*s" of Kafka's writing (Unsel 190): the dynamic of paradox-formalization-paradox.

WORKS CITED

- Anz, Thomas. *Franz Kafka*. 2nd ed. Munich: C. H. Beck, 1989. Print.
- Brod, Max. *Franz Kafka: Eine Biographie*. Frankfurt a.M.: Fischer, 1962. Print.
- Brune, Carlo. "Ein enterbter Sohn" Studie zu Franz Kafka's Brief an den Vater. Essen: Verlag Die Blaue Eule, 2000. Print.
- Deleuze, Gilles and Guattari, Felix. *Kafka: Towards a Minor Literature*. trans. Dana Polan, foreword Réda Bensmaïa. Minneapolis, MN: University of Minnesota Press, 1986. Print.
- Gibian, George. "Dichtung und Wahrheit. Three Versions of Reality in Franz Kafka." *German Quarterly* 30 (1957): 20-31. Print.
- Hasselblatt, Dieter. *Zauber und Logik: Eine Kafka-Studie*. Köln: Wissenschaft und Politik, 1964. Print.
- Honegger, Jürg Beat. *Das Phänomen der Angst bei Franz Kafka*. Berlin: E. Schmidt, 1975. Print.
- Kafka, Franz. "Der 'Brief an den Vater.'" In: *Franz Kafka Kritische Ausgabe: Nachgelassene Schriften und Fragmente II*. Hrsg. Jost Schillemeit. Redaktion Hans-Gerd Koch. Frankfurt a.M.: Fischer Taschenbuch Verlag, 2002. 143-217. Print.
- Müller-Seidel, Walter. "Franz Kafkas Brief an den Vater: Ein literarischer Text der Moderne." *Orbis Litterarum* 42 (1987): 353-374. Print.
- Pfeiffer, Joachim. "Ausweitung der Kampfzone: Kafkas Brief an den Vater." *Der Deutschunterricht* 52 (2000): 36-47. Print.
- Politzer, Heinz. "Franz Kafka's Letter to His Father." *Germanic Review* 28 (1953): 165-179. Print.
- — —. *Franz Kafka. Der Künstler*. Frankfurt a.M.: Fischer, 1965. 405-416. Print.
- Rattner, Josef. *Kafka und das Vater-Problem; ein Beitrag zum tiefenpsychologischen Problem der Kindererziehung. Interpretation von Kafkas "Brief an den Vater."* München: E. Reinhardt, 1964. Print.
- Schaefer, Margret. "Beyond Oedipal Narrative: Kafka and Contemporary Psycho-analysis." *Journal of the Kafka Society of America* 11 (1987): 52-60. Print.
- Sokel, Walter H. *The Myth of Power and the Self: Essays on Franz Kafka*. ed. Liliane Weissberg. Detroit: Wayne State University Press, 2002. Print.

- Stölzl, Christoph. "Brief an den Vater." In: *Kafka-Handbuch* in zwei Bänden. Hrsg.: Hartmut Binder. Stuttgart: Kröner, 1979. Print.
- Unsel, Joachim: "Nachwort zu Franz Kafka's 'Brief an den Vater.'" In: Kafka, Franz. *Brief an den Vater: Faksimile*. Hrsg. und Nachwort Joachim Unsel. Frankfurt a.M.: Fischer, 1994. Print.
- Wagenbach, Klaus. *Franz Kafka: in Selbstzeugnissen und Bilddokumenten*. Reinbek bei Hamburg: Rowohlt Verlag, 1964. Print.
- Weidner, Daniel. "3.3.2. <Brief an den Vater>." In: *Kafka-Handbuch: Leben – Werk – Wirkung*. Hrsg. Manfred Engel, Bernd Auerochs. Stuttgart, Weimar: Verlag J. B. Metzler, 2010. 293-301. Print.

CHAPTER 4: LACAN, FORMALIZATION, AND PSYCHOANALYSIS

LACAN AND THE PROBLEM OF FORMALIZATION⁴⁴

Rigor was a crucial concern for Jacques Lacan in his early and middle teachings. Drawing on the burgeoning disciplines of cybernetics and game theory, Lacan strove to transform psychoanalysis into a rigorous science by formalizing its technique. In his 1953 “The Function and Field of Speech and Language in Psychoanalysis,” he stresses that “[p]sychoanalysis can provide scientific foundations for its theory and technique only by adequately formalizing the essential dimensions of its experience, which ... are intersubjective logic and the temporality of the subject” (*FF Écrits* 298). Formalization was also a central goal for cybernetics in the 1950s, which studied and streamlined the transmission of information in communication systems. In the words of one of its pioneers, Claude Shannon, “semantic aspects of communication are irrelevant to the engineering problem. The significant aspect is that the actual message is one *selected from a set* of possible messages. The system must be designed to operate for each possible selection” (Shannon & Weaver 3). Cybernetics sought to provide a mathematized account of possible *selections* and *combinations* of information by relying on circuits, tables and functions.

This agenda enticed Lacan to draw on cybernetics in his efforts to turn psychoanalysis into a science. For Lacan, cybernetics was “the science of the combination of places as such,” (II.XXIII 299) a machinery in which the subject could variously appear or fail to appear. Indeed, as Lydia Liu demonstrates, Lacan conceived of the unconscious as a cybernetic

⁴⁴ This introduction and Part II of this chapter were originally published by the Edinburgh University Press as: Kuehne, Tobias. “Mathemes *avant la lettre*: Lacan’s Language Functions in ‘The Instance of the Letter’ as Triggers for Formalization.” *Paragraph* 40.2 (2017). 193-210. Print. Right to reprint retained by the author.

structure (Liu a).⁴⁵ Cybernetics promised to furnish the formal tools to make psychoanalysis a science of man (Elmer 216), allowing Lacan to confidently proclaim in his Rome Discourse that “the problem of the foundations that must assure our discipline its place among the sciences [is] a problem of formalization” (FF, *Écrits* 284).

Lacan had been interested in formalizing the processes of the unconscious even before cybernetics gained wide currency. For example, he had already given a descriptive account of the subject’s temporal logic of first hesitating and then hastening to conclusion in “Logical Time and the Assertion of Anticipated Certainty” of 1945 (*Écrits* 197-226). However, Lacan’s formalization efforts increased and intensified under the influence of Claude Lévi-Strauss and Roman Jakobson’s application of cybernetics to structural linguistics in the early 1950s (Gheogegan 112). In his second seminar (1954/55) (II.XV, II.XVI)⁴⁶ and his 1957 version of the ‘Seminar on the Purloined Letter’ as printed in *Écrits* (11-64), Lacan began to play games and draw charts and diagrams that laid out the itinerary of the thinking and speaking subject. While those formalisms seem opaque at first, Bruce Fink lucidly lays out their rules and operations in *The Lacanian Subject*.⁴⁷ Lacan provides genuine formalisms of the unconscious during his cybernetic phase.

This chapter traces Lacan’s changing approach to formalization by proceeding in two parts. First, this chapter will analyze how Lacan pushes formalization to its breaking point by exposing the impossibility of laying out a predictive formalism for simple intersubjective interactions. Hitting upon this paradox in the interactions of lower-case others brought Lacan

⁴⁵ In her book, *The Freudian Robot*, Liu shows how Lacan’s formalisms appear as outright adaptations of Shannon’s diagrams in “Mathematical Theory of Communication” (see Liu b 188-89).

⁴⁶ Lacan, “XV. Odd or even? Beyond intersubjectivity” and “XVI. *The Purloined Letter*,” II.

⁴⁷ Bruce Fink, *The Lacanian Subject: Between Language and Jouissance* (Princeton NJ: Princeton University Press, 1995). Cf. esp. chapter 2 and Appendix 1.

to theorize the *Other*. In its second part, this chapter will discuss Lacan's changing approach to formalization. Specifically, it identifies the metonymy and metaphor functions in "L'instance de la lettre" as a turning point in Lacan's thought. Even though Lacan insists we take the formulas literally, i.e. as mathematical functions, following his necessarily leads to a breakdown in understanding that forces us to go through the split subject's disrupted constitutive process of metonymy and metaphor. Taking the language functions literally (and failing) *engenders* the effect they merely seem to signify. Disruption is a catalyst for the split subject's constitutive, self-formalizing process, and its role is crucial for understanding how Lacan reconceptualized the relation between psychoanalysis and science. Lacan does not provide a formalism for us to master, but puts it into operation. The language functions are *mathemes avant la lettre*.

PART I: LACAN, GAME THEORY, AND THE PROJECT OF FORMALIZATION

1. Monkey See, Monkey Do

A few years ago, a study conducted by primate researchers from Kyoto University and game theory experts from Caltech University had chimpanzees and humans face off in the schoolyard game of "Odd or Even" (also known as Matching Pennies) (Martin et al.). The game is simple: player 1 chooses one of two options—odd or even (also expressible as plus or minus, right or left, heads or tails, etc.). Player 2 has to guess player 1's choice. If player 2 guesses correctly, she receives a marble from player 1 (one payoff unit); if player 2 guesses incorrectly, she has to give a marble to player 1.⁴⁸ The study showed that chimpanzees

⁴⁸ The game is zero-sum: the total of the payoffs distributed among the players does not change. What one player gains, the other one loses. Zero-sum games are closed systems.

consistently outperformed their human counterparts. Game theory predicts the behavior of rational agents. Did the researchers reveal that chimpanzees are more intelligent than humans?

In game theory, a game is “well-defined” if it specifies four elements: 1) the players, 2) the actions that are available to them at each moment in which they can make a decision, 3) the payoffs that each player can attain at the end of a decision chain, and 4) the extent of information available to each player (i.e., whether the players know who the other players are, their payoffs, and the decisions available to them) (Rasmusen 11). These four elements constitute the rules of the game (ibid.), since optimal play for players that aim to maximize their payoffs can be deduced from those four elements. In other words, players can deduce a complete plan of optimal action—a strategy—from those four elements.

For the game “Odd or Even,” game theory suggests that the best strategy for both players is to randomize their choices, i.e. play each of their choices with a probability of $\frac{1}{2}$. Over a long iteration of rounds, each player will play each choice equally many times, with the expected payoff for both at zero. This is the best each player can do, because as soon as it is known that one player plays a certain choice with a probability greater than $\frac{1}{2}$, the other player can adapt her strategy accordingly and extract an expected payoff greater than zero from her opponent. The players’ combination of strategies (choosing each outcome with a probability of $\frac{1}{2}$) is “stable” because neither player has an incentive to deviate from their strategy. Each strategy is a best reply to the respective other, which makes them a Nash Equilibrium (Binmore 18): rational (i.e. payoff-maximizing) play tends towards a Nash Equilibrium.

The matrix below visualizes this information. Each quadrant represents one possible outcome of the game, with the lower left number representing the row player’s payoff, and the upper right number representing the column player’s payoff.

		Player 2	
		Even ($q = \frac{1}{2}$)	Odd ($1 - q = \frac{1}{2}$)
Player 1	Even ($p = \frac{1}{2}$)	<div style="display: flex; justify-content: space-between;"> +1 -1 </div>	<div style="display: flex; justify-content: space-between;"> -1 +1 </div>
	Odd ($1 - p = \frac{1}{2}$)	<div style="display: flex; justify-content: space-between;"> -1 +1 </div>	<div style="display: flex; justify-content: space-between;"> +1 -1 </div>

Figure 1: "Odd or Even" Payoff Matrix

"Just play randomly." — Such a strategy is easier said than followed. The human players did much worse than their chimpanzee counterparts, even though the only information available was the history of the game itself. In addition, the human players were told that their opponents were human also players.⁴⁹ Yet, it was the chimpanzees who figured out the game and soon played in accordance with the predictions of game theory: their play matched with the Nash Equilibrium (Martin et al. 2f.). Their play tended toward randomized play, i.e. of playing without rhyme or reason; human play did not. Because the chimpanzees' play had no discernible pattern, they played more *rationally* than humans, according to game theory.

2. The Intersubjective Prison—and Its Dilemmas

Breakdowns of rationality of these and similar kinds are not uncommon in game theory, and have been discussed within the discipline as well. An eminent example is Nigel Howard's *Paradoxes of Rationality* (Howard). However, such intradisciplinary discussions do not address the self-undoing tendencies that inhere in game theoretical notions of rationality and

⁴⁹ All games were played through an interpolated machinic interface: each player, human and chimpanzee alike, pressed a button. The results of each round were revealed on a computer screen, with no direct interaction between players. The only information available was the history of the game itself, and, to the humans, the fact that the other player was a human.

symmetric distribution of knowledge.⁵⁰ I therefore propose to turn to a thinker who had his own take on game theory, and who provided his own reading of the “Odd or Even” game: Jacques Lacan. Lacan dedicated several sessions of his second seminar to exploring the reasoning of that game through the lens of Poe’s short story “The Purloined Letter,” (cf. esp. II.XV, XVI, and XXIII), as well as his “Seminar on ‘The Purloined Letter’” that grew out of those sessions and was published in his 1966 *Écrits*.

Lacan diagnoses “the dialectic of ‘even and odd’” in the structural dynamic of Poe’s story (*Écrits* 12). The young officer Dupin has mastered the game better than anyone else because he can identify with his opponents’ thinking. He has adopted the skill from a schoolboy who won every game of odd or even by imitating the facial expressions of his opponent in order access his intellect. Through this method of “egomiming” (II.XV, 180), the schoolboy could decide whether the other player had chosen to repeat or deviate from his previous choice of an odd or even number of marbles in his hand and guess correctly (II.XV, 179f.). Dupin, in other words, solves the mystery of the stolen letter that is hidden from everyone by being put in plain sight because he has mastered the key skill to winning at “Odd or Even”: intersubjectivity, the ability to think what the other thinks.

Lacan, however, does *not* read Poe’s story as an exposition of the structure of the efficacy of dyadic intersubjectivity. Instead, he turns the story inside out by reading it as a pointer to intersubjectivity’s self-defeating nature:

We are here faced with reasoning which raises a certain number of problems. At first glance, it is a matter of simple psychological penetration, a kind of egomiming. The subject adopts a mirror position, enabling him to guess the behaviour of his adversary. (II.XV, 180)

⁵⁰ Howard, too, offers more of an unfolding of the paradox of metarationality in showing that, in a game with n players, n metagames can be constructed, one from each player’s perspective. Each metagame has its own set of metagames, and so on ad infinitum.

If we have two subjects in a relation in which one is able to mime the other's thinking process in order to anticipate and take advantage of his choice in "Odd or Even," the other subject is *also* able to mime the first one's thinking process. Player 2 thinks what player 1 thinks that player 2 thinks and adapts her choice to that (cf. *Écrits* 58, II.XVI, 181), and so forth. Intersubjective thinking leads to an infinite regress. To escape this impasse of undecidability, the player choosing the number of marbles must play *unpredictably*, i.e. according to a law that is indiscernible, and without intention. His best chances lay in "acting like idiot" (Lacan, *Écrits* 58), or in *randomizing* his choices with equal probability (the game theoretical Nash Equilibrium strategy). The rational (payoff-maximizing) thing to do is to choose without rhyme or reason. Game theory proves it, theoretically and experimentally: the chimpanzees outdo the humans.

Matters get worse from there. To randomize means to choose completely by chance. However, psychoanalysis does not allow for random choices in the sense of their being undetermined (II.XXIII, 295). Indeed, Lacan emphasizes it is one of Freud's central convictions "that a number is never chosen at random" (*Écrits* 60). As a randomizing player in the game, I am bound to oscillate between two equally valid choices until—I find a criterion to choose by. Consciously or unconsciously, I always choose according to some law. Until then, I waver between the two choices. Once I have drawn on this law, it becomes potentially knowable, and it is that law that my opponent attempts to figure out, which I try to preempt, and so forth.

It may be objected at this point that Lacan's attack on intersubjectivity applies only to a small subset of interactions, namely those that are zero-sum. Due to the asymmetric payoff structure that makes one player lose what the other one wins, their interaction falls into interminable oscillation because each of them has to be one step ahead of the other, without there being an assurance that one can remain ahead of the other. In non-zero-sum games with

symmetric payoff structures, both players can collectively lose or win, depending on the setup of the game. Behaviors can be coordinated such that each individual player tries to maximize her own payoff, *regardless* of the other player's payoff.

The paradigmatic example is the prisoner's dilemma: two players can independently choose to cooperate with the other player, or to defect. If both players cooperate, each receives a payoff of 3. If one player cooperates and the other defects, the cooperator receives a payoff of zero, while the defector receives a payoff of 5. If both players choose to defect, each receives a payoff of 1.

		<u>Player 2</u>	
		Cooperate	Defect
<u>Player 1</u>	Co-operate	+3 +3	+5 0
	Defect	0 +5	+1 +1

Figure 2: Prisoner's Dilemma Payoff Matrix

The game theoretical solution to the prisoner's dilemma runs like this (with slightly different payoffs, cf. Howard 22): suppose you are player 1 (it does not matter which player you choose to be since the game's setup is entirely symmetrical). If player 2 cooperates with you, you can either also cooperate (and receive a payoff of 3) or defect (receive 5). Since your preferences are fully represented by the payoffs' number relations, and since you are a rational player (seeking to maximize your payoff), you would choose to defect if player 2 chooses to

cooperate. What if player 2 defects? If you cooperate, you receive 0; if you defect, you receive 1. A rational player would choose to defect in this scenario as well.

In both cases, the payoff for defection is higher if you defect than it is for cooperation. The strategy of defecting “strongly dominates” the strategy of cooperating. The player does not know the other player’s choice; nor does she need to know it, since, regardless of the other’s choice, she is better off defecting. The game theorist does not care about the other’s choice in the prisoner’s dilemma: it would not change the player’s decision.

This reasoning applies equally to player 2. As a result, under a game theoretical calculus, both choose to defect. Each player thus receives a payoff of 1—less than if both had chosen to cooperate. The game is a dilemma because rational play leads to a worse outcome than non-rational (non-maximizing) play. The literature proposing to “fix” the prisoner’s dilemma is vast and ranges from attempts to somehow change the payoff scheme – in which case it would no longer be the prisoner’s dilemma but a different game – to changing the argument by which it is evaluated – in which case it is no longer a game theoretical argument (cf. Howard 46ff.).

An attempt of the latter category was based on the claim that, since both players are symmetrically conceived and are payoff-maximizing, they must a) choose symmetrically (both make the same choice), and b) choose such that they maximize their payoffs. Hence, they both choose to cooperate. This argument has been debunked as the twins’ fallacy (cf. Binmore 9). Rationality results in both players making a (particular) symmetric choice, that of defect-defect (D-D). This symmetry does not entail, however, that rational players must choose symmetrically and, for that reason, make the particular symmetrical choice of cooperate-cooperate (C-C). *Because* players are rational, they choose symmetrically (D-D); the inference “because players choose symmetrically, the rational choice is C-C,” is a fallacious inversion,

which, moreover, strips each player of the rationality we assume for them in the first place: each would have been better off defecting, which again gets us into the bottom right corner of the matrix.

My analysis of the prisoner's dilemma is different from the twin's fallacy in that it does not concern itself with the choices of each player, but with the peculiarities of the reasoning process that must occur before the choice if we take all the assumptions of the game theoretical setup – especially that of the players' rationality – seriously and grasp their full implications. The problem that emerges is that of an oscillating, imaginary intersubjectivity that – like the game of “Odd or Even” – has to be interrupted by an outside intervention.

I start by assuming the role of player 1 in the prisoner's dilemma and by taking stock of what I know, all of which is strictly in keeping with the assumptions of game theory itself.

- 1): I know that I am playing against one opponent. Both of us have the choices of C and D at our disposal.
- 2): I know that my payoffs for C-C, C-D, D-C, and D-D are 3, 0, 5, and 1, respectively.
- 3): I want to maximize my payoff, i.e. I know that I am rational.⁵¹
- 4): I know that my opponent has the same payoff structure as I do.
- 5): By 3) and 4), I know that my opponent's rationality, that of maximizing payoffs, is isomorphic to mine. She has the same preferences as I do.
- 6): By 3) and 5), I know that my opponent knows that she is rational: if her rationality is isomorphic to mine, and I know that I am rational, she must know that she is rational.

⁵¹ In his seminal work on the iterated prisoner's dilemma, *The Evolution of Cooperation*, Robert Axelrod claims that a player need not assume rationality (reward maximization) for her opponent (18). However, if rationality is conceived as nothing but maximization of payoffs as they are captured by the numbers in the game, it becomes unclear what those numbers signify if players are not rational. In the last instance, one always chooses what one prefers, even if one prefers to punish oneself, in which case the highest payoff lies in depriving oneself of payoff. It is then this payoff of the second order that actually has to be captured in the game. Indeed, the payoff numbers and their relations already figure and organize a player's preference structure. In other words, rationality is not even an assumption that is added to assuming the knowledge of payoffs—the numbers already express rationality by making certain choices preferable over others qua their comparatively greater value.

7): By 5), 6), and 4), I know that my opponent knows that I am rational: if she knows that she is rational, and she knows that I have the same payoff structures as she does, she knows that I am rational.

This inference chain goes on infinitely: due to each player's self-knowledge and knowledge of symmetry of the game, each player knows about the other player what that player knows about himself/herself, and so on to infinity. These inferences all follow from the basic assumption of symmetry, rationality, and knowledge of the self as rational, all of which are so basic that no game theoretical version of the prisoner's dilemma can deny them.

As I play the game in my capacity as a rational player, I first reason: if player 2 cooperates, I want to defect, and if player 2 defects, I also want to defect. Hence, I want to defect. I know player 2 thinks the same about me, so we both want to defect. I would get a payoff of 1, and so would player 2. She knows that she would get a payoff of 1, and that I would get that same payoff. She knows that I know this.

I know that this is a lower payoff for me than if we both cooperated. She knows this too. And she knows that I know this. I therefore think (and so does player 2): it is more rational for us to both cooperate than to both defect. So now we both no longer want to defect, but to cooperate. But now it occurs to me that if player 2 cooperates with me, that I want to defect. I therefore now want to defect. Of course, player 2 knows this as well and now also wants to defect. Since we are perfectly rationally thinking beings (and inexhaustible at that), the reasoning above applies again, and we both want to cooperate, making us both want to defect, making us both want to cooperate . . .

The upshot of this is that both players' intersubjective knowledge of the other's thinking, isomorphic to their own, will always open a more preferable constellation of choices for them and plunge them into an oscillation between cooperation and defection that purely rational and perfectly intersubjective thinking cannot escape. Because, while I do not know

how the other player decides to act (and how could I, given the above?), I know how she thinks, because I think the same way. I try to take advantage of my knowledge of her thinking, she tries to do the same, I know that, try to take advantage of that, and so on and so on. Not only do I not know what my opponent decides to do – I do not know what she decides to do because I do know perfectly what she is thinking. Pure rationality and perfect intersubjectivity do not lead to the decision of C-C, as the twin's fallacy suggests—they never lead to *any* decision, since there is always a more attractive choice than the one currently arrived in the chain of reasoning.

The prisoner's dilemma is a real dilemma because one cannot even finish the game. The criterion for choosing defection over cooperation must therefore come through an external criterion beyond intersubjectivity, a third term (e.g. an assumption of general risk-averseness rather than risk-proneness). It is for this reason that game theory has to posit that we do not know what the opponent is doing. It has to cut short the intersubjectivity circuit by fiat. Intersubjectivity is not arbitrarily introduced—it inevitably emerges with the assumption of self-knowledge of rationality and symmetry of the game. It inheres, I argue, in the assumptions of the prisoner's dilemma itself. The real prisoner's dilemma is that its assumptions are set up such that they must contradict each other. This is all the more surprising since the conceptual origin of the prisoner's dilemma is rooted in an intersubjective question: “if two people can screw each other over or cooperate with one another, how will they behave?” Game theory cannot give the answer, as something outside the players' rationality must make that decision.

Game theory may stipulate the exclusion of intersubjective thinking. However, such a stipulation is arbitrarily imposed, blinds game theory to the dizzying thought processes that *do* occur in players' minds, and trades rationality's breakdown qua the interminability of its

reasoning for rationality's breakdown in outcomes that can sometimes betray their irrationality. Intersubjective thinking is a pernicious machine: it creeps in at unforeseeable moments (when symmetry occurs to us in a flash), and it is a process that cannot terminate by itself, but has to be terminated by a violent interruption. It is not restricted to games that are zero-sum. Moreover, intersubjectivity need not even occur between two subjects: the reasoning unfolded above in the case of the prisoner's dilemma could have unfolded against a completely imaginary opponent. Lacan indicates as much when he speaks of the imaginary dimension of intersubjectivity (e.g. I.XVIII, 224), our ineluctable urge to impute patterns to a machinic opponent in "Odd or Even," turning the game of even or odd into the game of guessing the machine's behavior (II.XXIII), and of the possibility of a lectern appearing as an ego (*Écrits* 421f.). Intersubjective thinking, with all its imaginary outgrowths, cannot simply be shed once and for all. Lacan therefore does not want to get "outside" but "beyond" intersubjectivity. The road to the beyond of a realm must be *through* it.

3. "To the Exit!" — Time, Exigency, and a Glimpse of the Other

Lacan indicates the road through and beyond intersubjectivity already in his 1945 essay "Logical Time and the Assertion of Anticipated Certainty," in which he presents his account of a very different prisoner's dilemma. Three prisoners receive the chance to win their freedom in a competitive game against each other. Each of them wears a disk that is either white or black. They know that the three disks distributed were chosen from a total of five disks—two black and three white. None of the prisoners can see their own disk, but they can see the other two prisoners' disks. In the given game, each prisoner wears a white disk; the two black disks remain unused. Whoever can *logically deduce* (there is no verbal communication or guessing) her color first has to make for the exit and his free (cf. *Écrits* 197f.). How do the logically wired prisoners find out their own color?

Lacan's logical solution to the game is rather quick, so it is useful to reconstruct it from the perspective of a prisoner (call her A), who again has to obtain the solution through intersubjective reasoning. A thinks the following (and all of the following thinking is strictly A 's!):

- 1) "If I saw two blacks, it would be immediately clear to me that I am a white.
- 2) However, I see B and C are whites. Since there are three whites and two blacks, I have to stay put and figure out: Am I a white or a black?
- 3) Suppose I am a black. What does prisoner B see and think in that case? He would think the following:
 - a) B : "I see A is black and C is white. Am I a white or a black?"
 - b) Suppose I am a black. What does prisoner C see and think in that case?
 - i) C : "I see A and B are blacks. I immediately deduce that I am a white. I head for the exit."⁵²
 - c) B : But I see that C does not head for the exit. I cannot be a black, so I am a white. I head for the exit."⁵³
- 4) A : But I see that B does not head for the exit. Since B has not made this inference by the time he should have, I cannot be a black, so I am a white. I head for the exit."⁵⁴

Of course, *each* of the three logical prisoners goes through this reasoning chain *simultaneously*.

According to the logical solution, all three thus leave at the same time.

At first glance, intersubjective reasoning seems to provide a clear and clean solution to this game. While none of the prisoners succeed in beating each other at the game itself, they all arrive at the correct solution through a chain of reasoning that is not only intersubjective but also *finite*. However, the logical smoothness of the above reasoning dissolves under the scrutiny of Lacan, who now "dons the philosopher's garb [of] the good logician, odious to the

⁵² This is the logico-temporal "zero-beat" of the reasoning chain.

⁵³ This is the logico-temporal "one-beat" of the reasoning chain.

⁵⁴ This is the logico-temporal "two-beat" of the reasoning chain.

world” (*Écrits* 199). He aims to outdo the logician in logical rigor and thus expose the game of the disks as the sophism – the logical trickery – that it is (*ibid.*).

Lacan cracks the formalism open by highlighting its crucially temporal dimension. Purely logical proofs should be a-temporal: while it may take a human (or even a computer algorithm) a certain amount of time to reason through a mathematical proof, the proof’s result always already inheres in the initial assumptions and rules of inference. The realm of pure logic does not know a “before” and an “after.” In the game of the disks, however, temporal delay – hesitation on the part of the players, their non-instantaneous acting – is *indispensable* to the deduction of the logical result itself. Because players are *not* moving right away, players conclude they are *not* blacks.

Far from being experiential data external to the logical process, the suspended motions are so necessary to it that only experience can make the logical process lack here the synchronicity implied by the suspended motions as produced by a purely logical subject. (*Écrits* 203, also cf. 204ff.)

Matters get worse from here, however, since in a triadic setup, the thought experiment’s reasoning process does not unfold in time without disruption. In a simple dyadic relationship (two players, one black and two white disks, with the two white disks being used in the game), the absence of instantaneous motion provides certainty for both that they are whites. In the triadic setup, however, there are two absences of instantaneous motion: In *A*’s reasoning process, there is a stage in which *C* does not head for the exit and another stage in which *B* does not head for the exit. There are two intervals requiring “the duration of a time of meditation” (*Écrits* 205), or two “logical beats.” Since *A* cannot rely on the simple binary of presence or absence of hesitation, the question arises: how long should *A* hesitate and meditate?

The answer: as little time as possible. For if *A* lets the other two players leave without running with them, *A* can never find out if she is a white or a black: the others could have left

because they saw she was a black, cutting the reasoning process short at 3c), or because they saw she was a white, arrived at 4). Time passes continuously between those two stages, so *A* has to seize the proper moment. Once *B* and *C* are gone, *A* cannot decide any more (*Écrits* 206). *A* is then out of the game of subject constitution.

Due to the indeterminacy of “how long” a logical beat is, a self-amplifying tension arises: the longer each player waits, the higher the certainty that they are a white; but since all certainty is lost once two players leave without the third, each of them also wants to spring into action and leave as quickly as possible. Lacan states:

It is thus not because of some dramatic contingency, the seriousness of the stakes, or the competitiveness of the game, that time presses; it is owing to [*sous*] the urgency of the logical movement that the subject precipitates both his judgment and his departure. (*Écrits* 206)

In a thought experiment in which the players are assumed to be homogenous in all relevant regards, all three players would of course still leave at the same time, although, paradoxically, it would become impossible to predict exactly when this would happen. Only one thing is clear: as the players are forced to precipitate their decisions, they leave before they have full logical certainty. The crucially temporal dimension of the thought experiment has led the logical formalism to unravel by itself. Logic has defeated logic.

The precipitated departure by all three players leads to a necessary interruption in their run for the exit. Since none of them can be fully certain that they are whites (none of them has seen two blacks, and it is not the case that any one of them has seen a white, a black, and an immediate departure by one of them; they have seen two whites and one continuous hesitation, which is not enough for full logical certainty), all of them call their precipitated conclusion into question. A second hesitation – simultaneously by all three – occurs. Since none of them continued to run, which would only have been possible if that non-stopping player had seen one black, one white, and an initial hesitation, all of them can now be certain

that they are whites and head for the exit. “The truth of the sophism thus only comes to be verified through its *presumption*, so to speak, in the assertion it continues. Its truth thus turns out to depend upon a tendency that aims at the truth” (*Écrits* 211).

Lacan’s thought experiment has led to a curious result: Even though every player went through an intersubjective reasoning process in the beginning, the temporal tension that emerged through the triadic arrangement of subjects sparked an effect that burst through and cut short that intersubjective reasoning process. A collective, reciprocal constitution of subjectivity emerges, not as an imaginary effect, but a truth effect, precisely because accuracy cannot be achieved through mere thinking without acting. The dimension of the game that unfolds beyond intersubjective thinking, i.e. the dimension in which “the form of *the other as such*” (*Écrits* 208) (the Big Other) emerges, unfolds between two moments of stumbling: that of stumbling into the running after the first hesitation, and that of stumbling out of the running before the second. What occurs before the first stumble is mere intersubjective thinking. What occurs after the second is the mutual objectification of subjects in the imaginary order: Everyone’s disk has been figured out, everyone can now run off on their own without worrying about the others. But in between, the game plays itself. Everyone is in the game, no one is a master over it. In a purely dyadic setup, this intermediate stage drops out: Intersubjective thinking immediately results in a jump into the imaginary. The intermediate stage introduces a qualitatively new dimension⁵⁵ to the game of subject constitution.⁵⁶

⁵⁵ A dimension whose importance to game theoretical thinking (“diplomatic strategy,” as he calls it) Lacan does not fail to point out at the end of the article (*Écrits* 212).

⁵⁶ This qualitative addition occurs in the transition from two to three players only. In an analogical setup with 4 players, three black disks, and four white disks, there would be two such intermediate stages, i.e. a replication, but not a qualitatively new addition. (Cf. fn.4, *Écrits* 213).

What is qualitatively different about this intermediate stage between those two stumbling moments in the game of subject constitution is that the symbolic order reveals itself without the distorting overlay of the imaginary. In a machinic coupling of bodies that are in touch with each other, the symbolic order plays itself out for a short while without the subjects falling into the trappings of the imaginary intersubjective relation.

4. Beyond Intersubjectivity: The Other, the Symbolic, Play

Reasoning through the game of the disks drives home the phenomenon that was already glimpsed in going through the analysis of “odd or even” and the prisoner’s dilemma, i.e. breaking through the confines of intersubjectivity to the third term that undergirds it. This insight, however, should not deceive us on Lacan’s fundamental stance on intersubjectivity: While it pulls the subject into the confusions and illusions of the imaginary, intersubjectivity itself is not an illusion that can be permanently dispensed with. Indeed, it is Lacan’s insistence on the subject’s “radical intersubjectivity” (I.XVII, 217) that separates him from the school of object relations. For Lacan, intersubjectivity precedes the image of subject (the integrated ego), which emerges “nachträglich, [grappling] with the supposedly original experiences . . . without ever leaving the domain of intersubjectivity” (ibid.). As the game of the disks (and the analysis of the prisoner’s dilemma) should have shown, the subject moves beyond intersubjectivity by going through it, not by shedding it.

The triadic structure breaks through the dyadic, because the former is already implicated in the latter: “There is never a simple duplicity of terms. It is not only that I see the other, I see him seeing me, which implicates the third term, namely that he knows that I see him” (I.XVII, 218). In order for you to see the other seeing you, it is a requirement that you be able to see seeing generally: in order to perceive the other as an other, one has to have a

conception of alterity as such, which cannot be seen from inside the dyadic reciprocal seeing. This is the third term beyond intersubjectivity, the truth “at the very foundation of intersubjectivity. It is situated where the subject can grasp nothing but the very subjectivity that constitutes an Other as an absolute” (*Écrits* 20).

This Other is precisely that with which the subject, unbeknown to itself, attempts to establish a relation: it always asks for “a quod on which the interrogation bears” (II.XV, 185). Although the Other is implicated in every imaginary dyadic intersubjective relation, it is also precisely that which eludes it: as an absolute alterity, it can never be turned into an image and thus appropriated by the subject. Qua inclusion of an image into an ego by the subject, the Other is that which is excluded (cf. II.XIX, 244.). The key is therefore to experience this Other (II.XIX, 246). To be in sync with it, the subject has to stumble through and out of the trappings of imaginary intersubjectivity. The subject does this through *playing*: “the intersubjective relation, which unfolds in the imaginary, is at the same time, in so far as it structures a human action, implicitly implicated in the rule of the game” (I.XVIII, 224). In the game of the disks, this could be seen in the action of running that happens between the two stumbles.

To access the Other is nothing but to be in sync with the symbolic order itself. The symbolic structures the subject’s universe as such (II.III, 29f.). It crucially participates in the process of subject constitution – “the subject follows the channels of the symbolic” (*Écrits* 30) –, and it does so in accordance with the dynamic of play.

By itself, the play of the symbol represents and organises, independently of the peculiarities of its human support, this something which is called a subject. The human doesn’t foment this game, he takes his place in it, and plays the role of the little pluses and minuses in it. (II.XVI, 192)

And again a few lines down:

From the start, and independently of any attachment to some supposedly causal bond, the symbol already plays, and produces by itself, its necessities,

its structures, its organisations. (II.XVI, 193. Also cf. I.XVII, 216f., and *Écrits* 434)

Understanding games and what it means to play is therefore the crux of understanding how the symbolic order, and thus the proper process of subject constitution, unfolds.

Against this background of trying to access the playful unfolding of the symbolic chain that structures a fundamentally intersubjective subject operating triadically rather than dyadically, Lacan introduces his model of the symbolic chain in the “Seminar on ‘The Purloined Letter.’” The game out of which he derives this chain is Fort/Da, the originary game through which the subject is born into language via an imaginary effect (*Écrits* 46). The game consists in the child’s reenactment of the contingent, potentially unending competition of absence versus presence (I.XVII, 218), symbolized in the contingent succession of pluses and minuses.

The binarism of (+) and (–) seems to suggest at first glance that the chain consists of only two terms. However, the crucial grouping is the triplet, as the important building blocs are three different kinds of triads of pluses and minuses: constancy (+++, – – –) (1), alternation (+ – +, – + –) (3), and dissymmetry (+ – –, – ++, ++ –, – – +) (2). The symbolization in triadic elements impose a self-restriction on how the chain can evolve: if the ultimate triplet of pluses and minuses constitute a (1), adding another plus or minus will never enable the new ultimate triplet thus created to be a (3)—and vice versa. A (1) can only be followed by a (1) or a (2), a (3) only by a (3) or a (2), while a (2) can be followed by any triplet. Similar restrictions emerge in the order of α , β , γ , and δ , which are the symbols for the different kinds of triplets of triplets that Lacan takes.⁵⁷ There would, in principle, be no stopping of

⁵⁷ Note that such restrictions could not occur if Lacan had grouped the pluses and minuses into pairs of symmetry (++ , – –) (1) and dissymmetry (+ – , – +) (2): after a (1), both a (1) and a (2) can follow. After a (2), a (1) and a (2) can follow as well. To begin to move, there must be more than two subjects.

creating ever new triplets of triplets, which would result in an ever greater memory of past events in the symbolic chain. Lacan remarks: “While the subject doesn’t think about it, the symbols continue to mount one another, to copulate, to proliferate, to fertilise each other, to jump on each other, to tear each other apart” (II.XV, 185).⁵⁸

To summarize: the symbolic chain is fundamentally predicated on autonomous self-structuration (that introduces restrictions despite contingent alternations in a binary scheme) and radical alterity. Both emerge out of its triadic make-up that both allow and require the symbolic chain to always include one element at the expense of excluding another, an operation that breaks impasses of symmetry and allows the subject to experience paradoxes (epistemological breakdowns) and stumble into something Other. After stumbling into the Other, the subject picks up steam again, runs the course of a new formalism, until the subject stumbles out of that again. The symbolic chain sets the subject on the alternating path of “paradox-formalism-paradox-formalism-...” All of this is hinted at in “Odd or Even,” the prisoner’s dilemma, and is more fully illustrated in the game of the disks.

PART II: LACAN’S SHIFTING APPROACH TO FORMALIZATION

Given his evolving understanding of formalization, the question for Lacan became how to operationalize it. Speaking about the playful unfolding of the symbolic chain is not the same as affecting that unfolding. This chapter takes “The Instance of the Letter in the Unconscious” as a crucial evolutionary point in Lacan’s thinking and approach. Play and stumbling become institutionalized by insisting on not taking them as signifiers, but to take them “literally.”

⁵⁸ For a detailed and accessible account of Lacan’s symbolic chain, cf. the Appendix I in Bruce Fink’s *The Lacanian Subject*.

1. Arresting the Attention

In “The Instance of the Letter in the Unconscious” (“L’instance de la lettre dans l’inconscient ou la raison depuis Freud” (*IL*, *Écrits* 493-528)), written and presented in 1957, four formulas arrest the attention:

$$\frac{S}{s}$$

Figure 3: Lacan’s
Saussurean algorithm
in *IL*

$$f(S)\frac{1}{s}$$

Figure 4: Lacan’s
development upon
the algorithm

$$f(S \dots S')S \cong S(-)s$$

Figure 5: The metonymy function

$$f\left(\frac{S'}{S}\right)S \cong S(+)s$$

Figure 6: The metaphor function

They disrupt the flow of the text and of Lacan’s speech, and it is not immediately clear whether they are mathematical expressions or mere jumbles of letters. It would seem that “L’instance” should continue Lacan’s trajectory of formalization. According to Roudinesco, Lacan became even more invested in formalization after his exposure to Jakobson’s essay ‘Two Aspects of Language and Two Types of Aphasic Disturbances’ in 1956, in which the cybernetic principles of selection and combination appear through the function of metaphor and metonymy (Roudinesco 271). Lacan takes up these two principles when he derives the metonymy and metaphor functions from the Saussurean algorithm S/s and declares: “It is the function of the subject, thus introduced, on which we must now dwell since it lies at the crux of our problem: (*IL*, *Écrits* 516).

How should one read the formulas? Given Lacan’s work leading up to “L’instance,” one should expect a formalism laying out the procedural unfolding of metonymy and metaphor. At the outset, Lacan asserts: “But how are we to take the letter here? Quite simply, literally [à la lettre]” (*IL*, *Écrits* 495). He calls S/s an algorithm, refers to the other three expressions as functions and presents them in mathematical notation. To take them literally would be to take them as *mathematical* algorithms and functions—which is to say, as mathematical machines that prescribe formal, computational procedures. Unlike his

formalisms from just a few years before, what those procedures consist in never becomes clear. Fink, who unpacked the cybernetic mechanism of the symbolic chain in the ‘Purloined Letter’ seminars, does not provide such an elucidation in his meticulous reading of “L’instance” in *Lacan to the Letter*, but tersely interprets the formulas as shorthand summaries for Lacan’s theory on metonymy and metaphor (Fink 98-105). Nancy and Lacoue-Labarthe’s seminal *The Title of the Letter* (a work on which Lacan remarks that he has “never been so well read” (XX 65)) professes to observe Lacan’s injunction to take the letter à la lettre by taking “‘algorithm’ in the strict sense,” (Nancy & Lacoue-Labarthe 34) and by uncovering the “determination of a calculative mode (and of a corresponding conceptuality) on the basis of which alone an object of science can be constructed” (ibid. 33). Just like Fink, however, the two authors never show how the algorithm operates or what it computes. Instead, they treat Lacan’s formulas as metaphors to be deciphered, attempting to deduce a systematic account of Lacan’s theory (for which they promptly incurred Lacan’s criticism [XX 66]).

The language functions in “L’instance de la lettre” are a stumbling block for (careful) readers of Lacan: precisely when his interest in the formalizing science of cybernetics appears to peak, his functions and algorithm cease to reveal a computational procedure. Has Lacan abandoned his projects of formalizing the subject’s entanglement in the structures of the unconscious and of leading psychoanalysis into the fold of the sciences? His abiding interest in mathematics throughout his career (he variously engages with set theory, graphs, operators of formal logic, topology and knot theory) suggests otherwise, and much insightful work in recent years has shown that Lacan did *not* abandon this question, but rather reconceived the relation between psychoanalysis and science in subtler and more complex terms (Milner, Ragland, Glynos & Stavrakakis, Hallward & Peden). Instead of molding psychoanalysis into a science as it was practiced in his day, Lacan asked how our conception of the sciences would

have to be *recast* in order to be able to accommodate psychoanalysis. This reconceptualization emerges, I argue, in the changing understanding of formalization that Lacan begins to develop in “L’instance de la lettre.”

2. Algorithmicity and a Blind Alley

Lacan enjoins us to take him literally, to take the letter à la lettre, which requires us to treat the Saussurean algorithm as a mathematical algorithm and the language functions as mathematical functions. Algorithms are the operative mechanisms that run inputs through a machine-like calculus to produce outputs. Similarly, functions relate inputs to outputs, either trivially by assigning numbers to each other in ordered pairs, or by a series of mathematical operators. Functions thus fall under the umbrella term ‘algorithm,’ and we need a criterion that allows us to determine whether or not an expression is an algorithm.

Neither functions nor algorithms have a universally agreed-upon definition in mathematics, despite their status as fundamental concepts (Ershov & Knuth 106). Nonetheless, the majority of mathematicians agrees that algorithms have the following characteristics (ibid.; also Rogers 2-3):

- 1) An algorithm carries out a *series of computational steps* (usually finite, although infinite series of computational steps are not automatically ruled out).
- 2) An algorithm is applied to *data* from its *domain*. An input datum is fed into the algorithm, which then yields an output datum (the result or solution).
- 3) Each computational step is *carried out mechanically and automatically*. It allows for no randomness, interpretation or ambiguity. All other things equal, a computational step yields the same output every time, as does the algorithm as a whole.
- 4) The *succession of computational steps is unambiguously determined*. After each computational step, there is either exactly one step to be carried out next, or none (in which case the algorithm comes to a stop). It is also known *which* step is to be carried out next (or if the algorithm stops). The decision about the next step is never left to the discretion of the one carrying out the algorithm.

Algorithms and functions are at the heart of cybernetics, as they prescribe automated computational processes. One basic example of an algorithm is a function such as $f(x) = x^2$. With the set of real numbers as its domain, the function's output is the square of the number that served as an input. These four conditions can be combined into the criterion of *algorithmicity*, to evaluate Lacan's algorithm and language functions in 'L'instance."

Applying this criterion quickly undoes any mathematical pretensions. Figures 1 and 2 are disqualified immediately because they are not formulated as anything resembling equations and, a fortiori, functions. The metonymy and metaphor functions also do not fulfil the conditions of algorithmicity. Condition 1) does not obtain because the (—) and (+) signs are not mathematical operators of subtraction and addition, but a visualization of Lacan's own notion of the signification bar which is crossed in the case of metaphor, but remains inviolate in the case of metonymy. As such, no computational steps can be taken. Condition 2) fails because, while the 'domain' of those functions could either be the set of signifiers or of letters, it is unclear how they could be treated as data and plugged into the left-hand side of the equation to yield outputs. Indeed, the formulas are not even equations, but only resemble them, as they feature congruence instead of equality signs. As such, automatic execution and unambiguous succession of computational steps (the third and fourth condition) also do not apply.

Taking Lacan literally leads down a blind alley (and a short one at that): the formulas do not deliver on their promise of an algorithmic operability. It seems as though Lacan is toying with us, and that his injunction is a playful irony. However, this does not authorize the reader to resort to one of two easy but misguided solutions that suggest themselves here: dismissing the formulas out of hand in frustration, or taking the license of treating them as metaphors, i.e. performing readings responding to questions such as, "What do they mean?"

or, “What lies behind them?”, or “Which part of Lacan’s theory do they capture?” Such approaches are tempting in that they either outright avoid the problem of how to engage with the formulas, or in that they assume the formulas to represent and substitute parts of Lacan’s thought. The former would be a cop-out, the latter is far from obvious.

Scholarly literature has consistently opted for one or both of these tacks, with varying results. Sokal and Bricmont take Lacan’s non-mathematicity as grounds for rejecting his teachings wholesale (Sokal & Bricmont). Marini claims mathematical ignorance to refrain from judgment (Marini 67-70). Gallop concludes that the functions’ ‘mathematical absurdity’ allow for a freely interpretive, even associative reading that compares Lacan’s and Jakobson’s theories of metonymy and metaphor without further dwelling on Lacan’s functions (Gallop 118-19), while Felman ignores the formulas altogether to liberate her interpretation of Lacan (Felman 94-207). Chaitin criticizes both of them and claims that, in “seeking to avoid metaphoric interpretation, [Gallop and Felman] end up being ensnared by it,” (Chaitin 1000) only to offer such a metaphoric interpretation of the formulas himself three pages later (ibid. 1003). Fink and Nancy and Lacoue-Labarthe, as discussed above, treat them as decipherable metaphors as well. Leupin, in turn, takes the formulas to be actual mathematical expressions, only to read his account of metonymy, metaphor, the Symbolic and the Real into the alleged operations of the formulas (Leupin). All of those accounts, whether they ignore or confront the formulas, launch into interpretations that provide the illusion of mastery over Lacan’s thought. How well those readings succeed depends on how well the authors have understood Lacan in the first place, without the formulas’ support. None of those readings engages with Lacan’s demand to take the letter à la lettre, and none of them show how the formulas enact an automated operation of any kind.

3. Reconstructing the Stakes of Lacan's Mathematics and Science

Lacan creates a dilemma: he demands we take the formulas literally, but this leads nowhere. Yet, dismissing or interpreting the formulas would amount to disregarding Lacan's injunction. Lacan frustrates the reader. He introduces doubt, hesitation and impatience. It is not at all clear what he means by taking the letter à la lettre.⁵⁹ This most intuitive of notions turns out to be the most opaque, as the letters themselves are stubbornly silent. A different kind of play seems to be at work than that of unfolding the rules of a formalism contained in a little machine. Before we can take mathematics in Lacan literally, i.e. on its own terms, we have to answer the question: what are those 'own terms' on which they have to be understood? Instead of ignoring or deciphering the formulas, we need to take a detour and return to them after grasping more clearly Lacan's project of formalizing psychoanalysis. Lacan himself suggests as much in his opening declaration of "L'instance," saying that he wishes to "leave the reader no other way out than the way in, which I prefer to be difficult" (*Écrits* 493).

"L'instance de la lettre" introduces the functions of language in a context connecting two major foci of Lacan's thought, namely the unconscious and the subject. Lacan introduces the Saussurean expression because "[t]he point is (...) to define the topography of the unconscious. I say that it is the very topography defined by the algorithm: S/s " (*IL*, *Écrits* 515). This is not surprising, since Lacan emphasizes that the unconscious is structured like a language (*FF*, *Écrits* 258) and that spoken language unfolds according to the functions of metonymy and metaphor (*IL*, *Écrits* 515; see also *SS*, *Écrits* 799; *PU*, *Écrits* 835). The passage on the language functions ends by thematizing the emergence of signification from metaphor for the subject and the importance of dwelling on "the function of the subject, thus introduced

⁵⁹ For the intricacies and entanglements of taking Lacan literally, refer to Cummings.

(...) since it lies at the crux of our problem” (*IL*, *Écrits* 516). The passage on the functions of language shifts focus from the unconscious to the subject within just a few lines.

Lacan specifies the subject as the *Cartesian* subject, which emerged ‘at the historical apex of the reflection on the conditions of science’ (*ibid.*). He elaborates: “the notion of the [Cartesian] subject is indispensable even to the workings of a science such as strategy in the modern sense, whose calculations exclude all ‘subjectivism’” (*ibid.*). Grasping Lacan’s account of the Cartesian subject is thus crucial for understanding what he considers to be the conditions for a discipline to properly qualify as a science. Only such an understanding will enable us to appreciate the full extent of the ‘Copernican revolution’ (*ibid.*) that Freud inaugurated, a revolution that decentered the subject permanently instead of substituting one concentric schema for another (*SS*, *Écrits* 797).

Reconstructing a complete account of Lacan’s Cartesian subject would take the scope of a book-length study (which keen readers of Lacan have already provided (cf. Fink, *The Lacanian Subject*; Milner; Chiesa). However, any exposition has to start from the observation that, according to Lacan, Descartes’ subject is fundamentally marked by a split. A key text is “Science and Truth,” which investigates that question, and begins by Lacan asking: “Shall I say that I established the status of the *subject* in psychoanalysis last year? I went so far as to develop a structure that accounts for the state of splitting [*refente*] or *Spaltung* where the psychoanalyst detects it in his praxis” (*ST*, *Écrits* 855) ⁶⁰ For Lacan, the process of subject constitution inexorably unfolds under the aegis of that split. One can thus speak of the split itself as constitutive for the subject. On those grounds, Lacan polemicized against ego

⁶⁰ Also, cf. ‘The Seminar on the Purloined Letter,’ *E* 55, *FF*, *E* 292, ‘On a Purpose,’ *E* 367 (henceforward *P*) and *PU*, *E* 835.

psychology and object-relation psychology for disregarding that division (*FF*, *Écrits* 242-43),⁶¹ sustaining instead ‘a mirage that is not even debated: the completeness of the subject,’⁶² a subject that is in truth a collection of images enlisted to fortify the ego⁶³ and ‘represents the center of all resistances to the treatment of symptoms.’⁶⁴ Any honest science must maintain this split, as only such a subject can qualify as a scientific subject at all: “the division of the subject [is] the fate of scientific man” (*P*, *Écrits* 367). In “Position of the Unconscious,” Lacan asserts “how fraudulent it is to extend [the Cartesian *cogito*’s] privilege to phenomena endowed with consciousness, in order to grant them a status. For science [in the Lacanian sense], the *cogito* marks, on the contrary, the break with every assurance conditioned by intuition” (*PU*, *Écrits* 831).

Lacan’s fundamental disagreement with his opponents therefore consists in his rejection that the subject could be formalized into a neat structure. Rather, the subject is precisely that element in science that principally *resists* formalization. Lacan thus implies that the sciences, with their empirical, objective, experimental approach, have taken a completely wrong turn after Descartes due to their misconceived notion of an integral, distantly observing scientific subject. In other words, the sciences are not scientific at all. Psychoanalysis can remedy this misunderstanding if it avoids the traps of ego psychology and object-relation psychology. Science has to account for the subject’s resistance to formalization, not become complicit in suturing it.

⁶¹ Also cf. ‘On a Question Prior to Any Possible Treatment of Psychosis,’ *E* 541. Henceforward *QP*.

⁶² Lacan, ‘On the Subject Who Is Finally in Question,’ *E* 232.

⁶³ *FF*, *E* 242-43, ‘On a Question,’ *E* 541 and ‘The Direction of the Treatment and the Principles of Its Power,’ *E* 590 (henceforward *DT*).

⁶⁴ Lacan, ‘Aggressiveness in Psychoanalysis,’ *E* 118.

The reason why the sciences since Descartes have gone astray in their attempt to suture the subject is that they are preoccupied with exact knowledge at the expense of effects of truth. They have labored to cover up the Cartesian doubt by establishing exact statements to constitute coherent bodies of knowledge instead of making palpable “what is disturbing about truth” (*SS, Écrits* 797). Psychoanalysis will not enter the field of science “without the object’s status in science as such being thereby modified” (*ST, Écrits* 863) and “establishing a new order of the sciences” (*FF, Écrits* 284): it will become a science by reversing the primacy of exact knowledge over effects of truth, i.e. by reconfiguring what science is. Lacan advocates for overturning the very paradigm under which the sciences of his time operate and calls for a recovery of a notion of truth that *should* have become the paradigm of science, had they not so fundamentally misunderstood the Cartesian subject and the doubt it entertains. Lacan conceives of the Cartesian subject as “an essential correlate of science” (*ST, Écrits* 856) by casting “our experienced division as subjects as a division between knowledge and truth” (*ibid.*).

To overcome this double division, the subject pursues its desire for knowledge by tracing a circuitous path determined by the law of the signifier. This law, Lacan claims, is captured by the operations of metonymy and metaphor, i.e. the combination and selection of signifiers. The subject employs metaphor and metonymy to alleviate the irremediable lack of the Mother from its constitutive split. We have thus circled back around to metonymy and metaphor, whose operations need to be elucidated further by drawing on “L’instance de la lettre.” Metonymy operates solely in the realm of the signifier without ever crossing the bar of signification. The unconscious combines signifiers based on the criterion of verbal contiguity: two signifiers connect to each other because they sound similar, are imagined “next to” each other, or because one is a part of the whole that is the other. Metonymy thus employs play of

words, puns and jokes.⁶⁵ Since the unconscious is structured like a language, desire articulates itself in the movement from one signifier to the next without ever satisfying desire by accessing a signified. Lacan calls metonymy “the vehicle of desire” (*PU*, *Écrits* 843), and announces at the end of “L’instance”: “desire *is* a metonymy” (*Écrits* 528).

In contrast to metonymy, metaphor does not connect signifiers but replaces them: “Metaphor is, quite radically speaking, the effect of the substitution of one signifier for another in a chain [of signifiers]” (*MS*, *Écrits* 890). The metonymic chain is replete with signifiers that have occulted other signifiers. Those substitutions create subject-constituting effects, as they are instances of crossing the bar of signification. The metonymic sliding of signifiers comes to a temporary halt in a button tie effect (“point de capiton”) when a signifier (arbitrarily) attaches to another signifier in a moment of sudden crystallization (*SS*, *Écrits* 805). Truth emerges as an *effect on the subject* when it discovers new speech, new memories, a new structuration of itself.⁶⁶

However, these crystallizing moments never provide the actual access to the signified that the subject desires. The subject only ever obtains other signifiers, the irremediable lack of the Mother is never fulfilled and desire never ceases. Indeed, the subject is born(e) into language in the moment when the thing becomes irretrievably lost to the signifier: “[T]he symbol first manifests itself [in the *Fort! Da!* game] as the killing of the thing, and this death results in the endless perpetuation of the subject’s desire” (*FF*, *Écrits* 319).⁶⁷ The signifying effect of metaphors is illusory in that the signified never arrives, but *true* in that it nonetheless

⁶⁵ Lacan himself engages in metonymy in *IL* by entitling the third part ‘La lettre, l’être, et l’autre’ (*Écrits* 523).

⁶⁶ Lacan, ‘The Signification of the Phallus,’ *Écrits* 689 (henceforward *SP*). Also *PU*, *Écrits* 830 and 835.

⁶⁷ Lacan states elsewhere: ‘It is the world of words that creates the world of things’ (*FF*, *Écrits* 276). Also cf. *SS*, *Écrits* 813 and *MS*, *Écrits* 892.

affects the split subject. Metaphors are therefore symptoms of those moments of redirection, which can manifest themselves in psychoses, neuroses or perversions. Lacan says about psychosis: “nowhere is the symptom more clearly articulated in the structure itself, assuming one knows how to read it” (*QP*, *Écrits* 537). While desire is a metonymy for Lacan, “the symptom *is* a metaphor” (*IL*, *Écrits* 528).

As the subject’s entrance into the law of the signifier coincides with its unbridgeable split that instates a desire that can never be fulfilled, desire becomes audible as an ever-unfolding chain of spoken signifiers. Metonymy and metaphor thus form an inextricable dyad that impels the interminable process of subject constitution. Both are ‘effects that are determinant in instituting the subject’ (*SP*, *Écrits* 689). “[A]s an effect of language, in that he is born of this early split, the subject translates a signifying synchrony into the primordial temporal pulsation that is the constitutive fading of his identification” (*PU*, *Écrits* 835). This effect must be one of a violent displacement, as it can only be achieved by metaphor’s illusory signifying effect. Lacan associates synchrony with metaphor and diachrony (“temporal pulsation”) with metonymy (*SS*, *Écrits* 805). Metaphor is the catalyst that constantly spurs the subject along the metonymic signifying chain. In Lacan’s words, “[t]his means that the most serious reality, and even the sole serious reality for man, if one considers its role in sustaining the metonymy of his desire, can only be retained in metaphor” (*MS*, *Écrits* 892).

Disruption and discontinuity therefore play a crucial role in Lacan’s conception of the subject’s entanglement with the signifying chain. Constitutive moments occur when a signifier intervenes, replaces another one and jolts the subject, whether in a button tie effect or in the creation of a symptom. It is not even possible to conceive of the subject without those experiences that Lacan calls cuts: “The cut made by the signifying chain is the only cut that verifies the structure of the subject as a discontinuity in the real” (*SS*, *Écrits* 801). The Lacanian

analyst has to account for the subject's fundamental split by bringing it to the fore during the analytic session: "we analysts must bring everything back to the cut qua function in discourse, the most significant being the cut that constitutes a bar between signifier and the signified. Here we come upon the subject who interests us" (ibid.).

In addition to his clinical views, Lacan espoused the same importance of disruption in his theoretical teachings. It was this emphasis on disruption that set him apart from psychoanalytic dogma at the time.⁶⁸ Disruptions are not glitches to be eliminated. The subject is constantly moving along the tracks of signification and can only be jolted and redirected, not permanently "fixed" or halted. There would be no process of subject constitution (and thus no subject at all) if the subject were not disturbed in its constitution. Discontinuity and subject constitution cannot be separated for Lacan (*SS*, *Écrits* 799-800).

This reconstruction of Lacan's conception of the subject, truth and disruption implies a notion of formalization that is much more intricate than the algorithmicity criteria. There is no smoothly unfolding process to be mastered and imposed on the subject. Rather, the subject always already formalizes. As Lacan puts it: "If we conduct the subject anywhere, it is to a deciphering which assumes that a sort of logic is already operative in the unconscious" (*SS*, *Écrits* 796). Neither the subject nor the analyst consciously know or control this process. Unlike his contemporaries in ego psychology, Lacan is not interested in bringing a stable, schematic structure of the subject or its constitutive process to consciousness, as such a structure is an illusion borrowed from the positive sciences. It is not the scientist (or the analyst) who formalizes at her leisure, but the subject, whose very being is a constant process of formalization. This process can at best be redirected through irritation from the outside

⁶⁸ His expulsion from the IPA in 1953 was on the grounds that he distressed his analysands with variable, unpredictably timed sessions. Cf. *JL* 386-87.

through a carefully timed intervention whose effects are nonetheless unpredictable: “The whole psychoanalytic tradition supports that the analyst’s voice can intervene only if it enters at the right place, and that if it comes too early it merely produces a closing up” (ibid.). Lacan disavows the illusion that formalization involving the subject could somehow be mastered, precisely because it is always already operative. It can only be nudged into operation when it is irritated and disrupted.

4. Mathemes Avant la Lettre and a Propulsive Effect

“L’instance de la lettre” of 1957 thus provides a vastly evolved notion of formalization from the ‘Seminar on “The Purloined Letter,”’ which Lacan finalized for publication in *Écrits* in the same year. The ‘Purloined Letter’ seminar featured an intricate mechanism of rules, the most sophisticated machine Lacan had yet devised. By contrast, the language functions act as triggers to precipitate the subject onto the path of formalization against its own conscious knowledge. The functions are turning points in Lacan’s thought.

The characteristic of introducing an impasse in signification by refusing to lay out a formal procedure to follow make the language functions the conceptual seed of a Lacanian notion he did not develop until fourteen years after “L’instance”: the *matheme*.⁶⁹ For Lacan, the *matheme* “alone is capable of being integrally transmitted” (XX 119) and “the only teachable discourse,” (LE 28) precisely because *mathemes* “are transcribed as dead-ends by the mathematizable, that is, the teachable in the Real” (LE 35). Just as the language functions, the *matheme* stages the signified’s retreat from the signifier in its resistance to being interpreted. The subject experiences its relation to the Real in this disruption of signification

⁶⁹ According to Bernard Burgoyne, Lacan introduced the *matheme* in December of 1971. Burgoyne 81.

and glimpses the process of its own constitution (Charraud 110-13). Douglas Aoki fleshes this out by showing how the mathemes, through their resistance to being read, do not *explicate* how the subject comes into contact with the Real, but make the subject *experience* this contact: Lacan's "mathemes dramatize the theory they situate" (Aoki 13). The mathemes operate the way they do because they look mathematical, while not in fact being mathematical.

The language functions perform the same gesture. Outwardly, they display the veneer of mathematics. In truth, however, they already enact the mathemes' crisis of reading: they draw attention to themselves, but instead of delivering (the illusion of) a signified, they empty themselves out and insist on being mere letters. While the language functions do not yet appear in the context of Lacan's later exposition of the Name-of-the-Father, *jouissance* and the absence of the sexual relation, they already operate as mathemes "avant la lettre," a hinge between Lacan's cybernetic phase and the post-cybernetic mathemes.

Of course, this essay exhausts neither the richness of Lacan's use of mathematics writ large nor the complexity of his notion of formalization. His schemas, the graphs of desire, his topological figures and the Borromean knots (to name just a few), all operate differently in the particular contexts in which they appear and have to be approached accordingly. While such a study would go beyond the scope of this essay, there are promising routes of investigation for each of them, as each promise to highlight new facets of Lacan's grappling with the problem of formalization. With the schemas, it would be imperative to follow their successive transformations and distortions to understand the "excess endemic to any formalization that is presented in the intuitive [imaginary] realm" (*QP*, *Écrits* 571). Lacan's schemas, which begin to appear in *Seminar II* (cf. *II* 109, *II* 243), likely aim for a breakdown in neat imaginary formalizations, a phenomenon he investigated when studying the Freudian ego and its entanglement in the imaginary. The key to the graphs of desire seems to be their self-

reproducing, expansive and permutative dynamic that self-awarely crystallizes various button ties in a formalization process. Lacan emphasizes that he does not provide a meta-language, but gives “indices of an absolute signification” that is constantly unfolding (*SS*, *Écrits* 816): he shows a formalization in progress. His treatment of the Möbius strip, in turn, demonstrates the subject’s overcoming of an impasse in formalizations by shedding its lack of an outside by inflating into a torus, while the Borromean knot presents itself as a riddle to be solved and openly invites the subject to play, permute and resolve its tangles. All of these figures stress different aspects of Lacan’s complex understanding of formalization, and none of them are fully congruent with the cybernetic machines’ layout of rules or the language functions’ nudging the subject into formalization under the law of the signifier.

However, instead of getting lost in the intricacies of Lacan’s various tacks on formalization, it is worth dwelling a little more on the language formulas. Lacan’s “Metaphor of the Subject,” which takes up and expands upon his language functions (*Écrits* 890), closes by countering the misconception that mathematical discourse, because of its clarity, is exempt from rhetorical effects: “Doesn’t this provide the status of rhetorical effects, in showing that they extend to all signification? Let people object that they stop at mathematical discourse—I will agree all the more in that I place the highest value on that discourse precisely because it signifies nothing” (*MS*, *Écrits* 892). According to scientists, mathematics can be taken literally because it does not equivocate: it is exact. Lacan responds by pointing out that this is so *not* because mathematical discourse unambiguously refers to signifieds without any room for error, but because it does not refer to any signifieds *at all*. Mathematical discourse, in its algorithmic, exactly determined operationality, unfolds purely in the realm of the signifier and has indisputable rules of connecting and replacing them. Mathematics is thus exemplary for Lacan in two ways: its unreadability (i.e. irretrievability of signifieds) and its covering up of this

unreadability by providing a smooth texture of operation. Mathematics exemplifies the subject's metonymic and metaphoric itinerary of signification in that it operates purely on the plane of the signifier *and* is blind to that enclosure. Lacan aims to lay bare both of those gestures by disrupting their continuous flow, i.e. by revealing mathematical discourse in its status as a naked letter. Mathematics' unreadability provides a purified demonstration of the subject's entanglement in signification and its illusions, provided one can be jolted into sensing this unreadability.

Having reconstructed the particularities of formalization that are at stake in the language formulas, and having become aware of the effects they are supposed to create, the question becomes: *do* they create that effect? *Do* the language formulas propel a subject onto metonymic and metaphoric tracks of signification, or do they fail at the outset and never become operative? Looking back at our own stumbling over the letters, our frustration with, departure from, and return to them, the answer is yes: we experienced a disruption in signification. Our subsequent insistence on taking the letters *à la lettre* introduced doubt, hesitation and impatience, ultimately impelling us on a path of signification. In taking that path, we have thus not only reconstructed Lacan's account of the split subject, the experience of truth as an effect and the role of metaphor and metonymy in it—we have experienced the effect of subject constitution. We have linked together signifiers *because* we are split subjects: taking the position of the Cartesian cogito, we demanded exactitude, which vanished in the face of the recalcitrant, silent, frustrating letters since we did not know how to take them literally. Lacan says in "L'instance": "[i]t wasn't going very far to say the words with which I momentarily dumbfounded my audience: I am thinking where I am not, therefore I am where I am not thinking" (*LI, Écrits* 517). The place of our thinking was not in what we enunciated, but in the fact that we were first confounded and then precipitated into speech. The language

functions had the effect of leading us into and through this experience of the subject. It is now, at this button tie moment, that the signification chain comes to a close and threatens to devolve into an imaginary illusion of mastery over Lacan. However, our experience of reading Lacan should at least open us up to being disrupted and thrown off guard by him, without avoiding those stumbling blocks or facilely interpreting them away.

Does our approach place us above and beyond all the critics whom we have accused of avoiding or interpreting the language formulas? Have Fink, Gallop, Felman, Chaitin, Nancy and Lacoue-Labarthe, Leupin, and Sokal and Bricmont all gotten it “wrong,” while we “finally got it right”? Putting the question in this way would be preposterous, and would it fall back into Lacan’s discarded regime of truth as exactitude. On the one hand, we have done what (almost) all of the aforementioned critics have done. We have derived some new insights, namely about Lacan’s complex notion of formalization, about his evolving understanding of the relation between psychoanalysis and science and we have uncovered a connection between the language functions and the later mathemes, making “L’instance” a key text that pivots away from more straightforward notions of formalization Lacan was still espousing in other texts around that time. On the other hand, we have circumvented the double trap of avoiding or interpreting the language functions. Instead, we retrieved the place in which they appear and come operate.

Lastly, I do believe to have achieved something genuinely new in having traced an effect of truth in this essay, even if I can only re-tell it at this point: we have not just explicated a theory of a split subject that avoidingly retreats and interpretively advances—we have also witnessed ourselves making those retreats and advances. Indeed, every critic was propelled onto their own (more or less plausible) path by the letters, as were we. The recognition of that effect is something that is only happening now. This may seem like a minute difference hardly

worth mentioning. Yet from a Lacanian perspective, it makes all the difference. The former way of reading Lacan is based on a notion of science that he disavows; the latter effects the Lacanian understanding of science by putting it into operation.

WORKS CITED

- Aoki, Douglas Sadao. "Letters from Lacan," *Paragraph* 29:3 (2006). 1-20. Print.
- Axelrod, Robert. *The Evolution of Cooperation*. New York, NY: Basic Books, Inc., 1984. Print.
- Binmore, Ken. *Playing for Real: A Text on Game Theory*. Oxford: Oxford University Press, 2007. Print.
- Bowie, Malcom. "Language and the Unconscious." In: Bowie, *Lacan* (Cambridge, MA: Harvard University Press, 1991), 44-87. Print.
- Burgoyne, Bernard. "From the letter to the matheme: Lacan's scientific methods," in *The Cambridge Companion to Lacan*, ed. Jean-Michel Rabaté (Cambridge, UK: Cambridge University Press, 2003), 69-85. Print.
- Chaitin, Gilbert D. "Lacan's Letter," *Modern Language Notes* 103:5 (1988), 995-1011.
- Charraud, Nathalie. "Mathème," trans. Dominique Hecq, in: Huguette Glowinski, Zita M. Marks and Sara Murphy (eds.), *A Compendium of Lacanian Terms* (London: Free Association Books, 2001). Print.
- Chiesa, Lorenzo. *Subjectivity and Otherness: A Philosophical Reading of Lacan* (Cambridge, MA: MIT Press, 2007). Print.
- Cummings, Brian. "Literally Speaking, or, the Literal Sense from Augustine to Lacan," *Paragraph* 21:2 (1998), 200-226. Print.
- Elmer, Jonathan. "Blinded Me with Science: Motifs of Observation and Temporality in Lacan and Luhmann," in William Rasch, Carry Wolfe, *Observing Complexity: Systems Theory and Postmodernity* (Minneapolis, MN: The University of Minnesota Press, 2000), 215-246. Print.
- Ershov, A.P. & Knuth, D.E. *Algorithms in Modern Mathematics and Computer Science* (Berlin, Heidelberg, New York: Springer-Verlag, 1981). Print.
- Felman, Shoshana. "Turning the Screw of Interpretation," *Yale French Studies* 55/56 (1977). Print.
- Fink, Bruce. *The Lacanian Subject: Between Language and Jouissance*. Princeton, NJ: Princeton University Press, 1996. Print.
- — —. "Reading 'The Instance of the Letter in the Unconscious'" in *Lacan to the Letter: Reading Écrits Closely* (Minneapolis, MN: University of Minnesota Press, 2004). Print.
- Gallop, Jane. *Reading Lacan* (Ithaca, NY: Cornell University Press, 1985). Print.

- Gheogegan, Bernard Dionysius, "From Information Theory to French Theory: Jakobson, Lévi-Strauss, and the Cybernetic Apparatus," *Critical Inquiry* 38:1 (2011), 96-126. Print.
- Glynos, Jason & Stavrakakis, Yannis, eds., *Lacan & Science* (Karnac, 2002). Print.
- Hallward, Peter & Peden, Knox, eds. *Concept and Form*, vols. 1 and 2 (London: Verso, 2012). Print.
- Howard, Nigel. *Paradoxes of Rationality: Theory of Metagames and Political Behavior*. Cambridge, MA, London UK: The MIT Press, 1971. Print.
- Jakobson, Roman. "Two Aspects of Language and Two Types of Aphasic Disturbances,; in Roman Jakobson and Morris Halle, *Fundamentals of Language* (The Hague: Mouton & Co, 1956), 53-82. Print.
- Lacan, Jacques. "Seminar on "The Purloined Letter."" *Écrits*. trans. Bruce Fink, in collaboration with Heloïse Fink and Russell Grigg. New York, NY: W. W. Norton & Co., 2006. 11-61. Print.
- — —. "Logical Time and the Assertion of Anticipated Certainty." *Écrits*. 197-213. Print.
- — —. [FF] "The Function and Field of Speech and Language in Psychoanalysis." *Écrits*. 289. Print.
- — —. "The Freudian Thing, or the Meaning of the Return to Freud in Psychoanalysis." *Écrits*. 401-436. Print.
- — —. "The Instance of the Letter in the Unconscious," *Écrits* 493-528. *IL*. Print.
- — —. [SS] "The Subversion of the Subject and Dialectic of Desire," *Écrits* 799. Print.
- — —. [PU] "Position of the Unconscious," *Écrits* 835. Print.
- — —. [ST] "Science and Truth," *Écrits* 855. Print.
- — —. [MS] "Appendix II: Metaphor of the Subject," *Écrits* 890. Print.
- — —. [I.XVII] "XVII. The object relation and the intersubjective relation." The Seminar of Jacques Lacan. Book I: Freud's Papers on Technique 1953-1954. trans., notes John Forrester. New York, NY: W. W. Norton & Co., 1991. 209-219. Print.
- — —. [I.XVIII] "XVIII. The symbolic order." *Seminar I*. 220-233. Print.
- — —. [II.III]. "III. The symbolic universe." The Seminar of Jacques Lacan. Book II: The Ego in Freud's Theory and in the Technique of Psychoanalysis. ed. Jacques-Alain Miller, trans. Sylvana Tomaselli, notes John Forrester. New York, NY: W. W. Norton & Co., 1991. 27-39. Print.

- — —. [II.XV] “XV. Odd or even? Beyond intersubjectivity.” *Seminar II*. 175-190. Print.
- — —. [II.XVI] “XVI. The Purloined Letter.” *Seminar II*. 191-205. Print.
- — —. [II.XIX] “XIX. Introduction of the big Other.” *Seminar II*. 235-247. Print.
- — —. [II.XXIII] “XXIII. Psychoanalysis and cybernetics, or on the nature of language.” *Seminar II*. 294-308. Print.
- — —. [XX] *The Seminar of Jacques Lacan, Book XX (Encore 1972-1973)*, trans., notes Bruce Fink (New York, NY: W. W. Norton & Co, 1998). Print.
- — —. [LE] “L’Étourdit,” *Scilicet* 4 (1973). Print.
- Leupin, Alexandre. *Lacan Today: Psychoanalysis, Science, Religion* (New York: Other Press, 2004). Print.
- Liu, Lydia H. [Liu a] “The Cybernetic Unconscious: Rethinking Lacan, Poe, and French Theory,” *Critical Inquiry* 36:2 (2010). 288-320. Print.
- — —. [Liu b] *The Freudian Robot. Digital Media and the Future of the Unconscious* (Chicago, IL: The University of Chicago Press, 2010). Print.
- Marini, Marcelle. *Jacques Lacan: The French Context*. trans. Anne Tomiche (New Brunswick, NJ: Rutgers University Press, 1992). Print.
- Martin, Christopher Flynn, et al. “Chimpanzee choice rates in competitive games match equilibrium in game theory predictions.” *Nature. Scientific Reports* 4. Art number: 5182. DOI: 10.1038/srep05182. Received 11 April 2013, Accepted 13 May 2014, Published 05 June 2014.
<http://www.nature.com/srep/2014/140605/srep05182/full/srep05182.html> (Last accessed March 3, 2021). Web.
- Milner, Jean Claude, *L’œuvre Claire: Lacan, la science, la philosophie* (Paris: Seuil, 1995). Print.
- Nancy, Jean-Luc & Lacoue-Labarthe, Philippe. *The Title of the Letter: A Reading of Lacan*. trans. François Raffoul and David Pettigrew (Albany, NY: State University of New York Press, 1992). Print.
- Plato. *Laws*. Loeb Classical Library. trans. R. G. Bury. Cambridge, MA: Harvard University Press, 1961. Print.
- Ragland, Ellie, ed. *Critical Essays on Jacques Lacan* (New York, NY: G.K. Hall, 1999).
- Rasmusen, Eric. *Games and Information: An Introduction to Game Theory*. 4th ed. Wiley-Blackwell, 2006. Print.
- Rogers, Hartley. *Theory of Recursive Functions and Effective Computability* (McGraw-Hill, Inc., 1967). Print.

Roudinesco, Elisabeth. *Jacques Lacan*. trans. Barbara Bray (New York, NY: Columbia University Press, 1997). Print.

Shannon, Claude E. & Weaver, Warren. *The Mathematical Theory of Communication* (Urbana, IL: The University of Illinois Press, 1949). Print.

Sokal, Alan & Bricmont, Jean. *Intellectual Impostures: Postmodern Philosophers' Abuse of Science* (London, UK: Profile Books, 1998). Print.

CHAPTER 5: FORGETTING *MARBURY*'S LESSON: A HISTORY OF QUALIFIED IMMUNITY

The history of qualified immunity has not yet been fully written. Although scholars have devoted countless pages to the doctrine, they have not paid enough attention to its origins, its metamorphoses, or the legal, political, and institutional pressures to which it responded. This chapter provides that missing history by deriving an analytical framework from *Marbury v. Madison*. *Marbury* was a suit against a government official in which the Supreme Court insisted on broad availability of individual rights, but withheld a remedy. In striking this middle ground, the *Marbury* Court eschewed political pressures, enhanced the judiciary's legitimacy and independence, and created a new judicial power. Qualified immunity, at its origin, struck a middle ground between a rigid formalism and judicial flexibility.

But qualified immunity is the story of a *Marburian* move gone wrong. When qualified immunity emerged as a defense in the waning years of the Warren Court, it hewed closely to *Marbury*'s wisdom of providing some formal structure, but maintaining enough indeterminacy to give judges room to maneuver. Under the stewardship of Justice Byron White, the Court brought that wisdom into an objective and uniform standard in *Harlow v. Fitzgerald*. But as the Court entered the Rehnquist years, it abandoned its *Marburian* middle ground, both retreating from its firm stance on rights and removing *Harlow*'s remedy from reach. Qualified immunity's history as a *Marburian* move gone wrong has upshots for our understanding of judicial power: not merely an authority to decide cases, judicial power seeks to carve out the threshold authority of deciding whether to decide. The authority of deciding to decide lies at the heart of judiciary's institutional concern for independence. The history of qualified immunity as an ever-rigidifying formalism shows that courts exercising this authority must retain a careful balance between recognizing individual rights and judiciously granting remedies.

INTRODUCTION

The history of qualified immunity has never been written. In the vast academic literature on the subject, its origins and development are usually glossed over, if they are mentioned at all.⁷⁰ As calls for qualified immunity’s overhaul mount, and the Supreme Court refuses to reconsider the doctrine,⁷¹ it is imperative that we understand how we got to this point. This chapter provides the missing history of qualified immunity.

1. The Current State of Qualified Immunity Doctrine

Qualified immunity creates a high bar for plaintiffs who seek money damages from government officials for federal rights violations under 42 U.S.C. § 1983⁷² or *Bivens v. Six Unknown Named Agents*.⁷³ Under *Harlow v. Fitzgerald*, “government officials performing discretionary functions” receive immunity for those violations if their “conduct does not violate clearly established statutory or constitutional rights of which a reasonable person would have known.”⁷⁴ The doctrine uses a functional approach: if the government official’s

⁷⁰ For some brief discussions of qualified immunity’s history, see William Baude, *Is Qualified Immunity Unlawful?*, 106 CALIF. L. REV. 45, 50-61 (2018); John C. Jeffries, *The Liability Rule for Constitutional Torts*, 99 VA. L. REV. 207, 250-52 (2013); Richard H. Fallon, Jr., *Bidding Farewell to Constitutional Torts*, 107 CALIF. L. REV. 933, 941-50 (2019).

⁷¹ In June of 2020, the Supreme Court denied certiorari in thirteen qualified-immunity cases, three of which petitioned for a recalibration of the doctrine. Jay Schweikert, *The Supreme Court’s Dereliction of Duty on Qualified Immunity*, CATO AT LIBERTY (June 15, 2020, 11:27AM), <https://www.cato.org/blog/supreme-courts-dereliction-duty-qualified-immunity>.

⁷² State officials are sued under § 1983. The statute in its current form reads: “Every person who, under color of any statute, ordinance, regulation, custom, or usage, of any State or Territory or the District of Columbia, subjects, or causes to be subjected, any citizen of the United States or other person within the jurisdiction thereof to the deprivation of any rights, privileges, or immunities secured by the Constitution and laws, shall be liable to the party injured in an action at law, suit in equity, or other proper proceeding for redress” 42 U.S.C. § 1983 (2020).

⁷³ 403 U.S. 388 (1971). Federal officials are sued under *Bivens*, albeit only for a limited number of constitutional violations. In *Ziglar v. Abbasi*, 137 S. Ct. 1843 (2017), the Court has recently emphasized that, in light of its *Bivens* jurisprudence in the last few decades, extending causes of action against federal officials to new contexts has become a “disfavored judicial activity.” *Id.* at 1857 (internal quotation marks and citation omitted).

⁷⁴ 457 U.S. 800, 817-18 (1982).

challenged acts were executive in nature, she can invoke qualified immunity; if the challenged acts are legislative, judicial, or quasi-judicial in nature, she receives absolute immunity for those acts.⁷⁵

Since the Court handed down *Harlow* in 1982, it has emphasized that rights must be defined at a high level of specificity in order to be considered “clearly established.”⁷⁶ Caselaw has to give “fair notice”⁷⁷ so that individual officers can “determine how the relevant legal doctrine . . . will apply to the factual situation[s that they] confront[]” on the job.⁷⁸ While the standard “do[es] not require a case directly on point,” the Court insists that “existing precedent must have placed the statutory or constitutional question beyond debate.”⁷⁹ Because qualified immunity was intended to “protect[] ‘all but the plainly incompetent or those who knowingly violate the law,’”⁸⁰ the Court has called it “an exacting standard.”⁸¹

⁷⁵ See, e.g., *Imbler v. Pachtman*, 424 U.S. 409 (1976) (granting absolute immunity for acts performed in a judicial function); *Burns v. Reed*, 500 U.S. 478, 484-87 (1991) (laying out the analytical framework); *Butz v. Economou*, 438 U.S. 478, 504-16 (1978) (same). See also *Buckley v. Fitzsimmons*, 509 U.S. 259, 273 (1993) (observing that qualified immunity “represents the norm” (quoting *Harlow v. Fitzgerald*, 457 U.S. 800, 807 (1982))).

⁷⁶ See, e.g., *Anderson v. Creighton*, 483 U.S. 635, 640 (1987) (holding that “the contours of the right must be sufficiently clear that a reasonable official would understand that what he is doing violates that right”); see also *City of Escondido v. Emmons*, 139 S. Ct. 500, 503 (2019) (“Under our cases, the clearly established right must be defined with specificity. ‘This Court has repeatedly told courts . . . not to define clearly established law at a high level of generality.’ (quoting *Kisela v. Hughes*, 138 S. Ct. 1148, 1152 (2018) (quoting *City of San Francisco v. Sheehan*, 135 S. Ct. 1765, 1775-76 (2015) (quoting *Ashcroft v. al-Kidd*, 563 U.S. 731, 742 (2011) (further citations omitted)))). See also Kit Kinports, *The Supreme Court’s Quiet Expansion of Qualified Immunity*, 100 MINN. L. REV. HEADNOTES 62 (2016) (discussing the Court’s progressive tightening of the standard).

⁷⁷ *Kisela v. Hughes*, 138 S. Ct. 1148, 1152 (2018) (quoting *Brosseau v. Haugen*, 543 U.S. 194, 198 (2004)). See also *Hope v. Pelzer*, 536 U.S. 730, 739 (2002).

⁷⁸ *Kisela v. Hughes*, 138 S. Ct. 1148, 1153 (2018) (quoting *Mullenix v. Luna*, 136 S. Ct. 305, 308 (2015) (quoting *Saucier v. Katz*, 533 U.S. 194, 201 (2001))). See also *City of San Francisco v. Sheehan*, 135 S. Ct. 1765, 1774 (2015) (quoting *Plumhoff v. Rickard*, 572 U.S. 765, 778-79 (2014)).

⁷⁹ *White v. Pauly*, 137 S. Ct. 548, 551 (2017) (quoting *Mullenix*, 136 S. Ct. at 308 (quoting *Ashcroft v. al-Kidd*, 563 U.S. 731, 741 (2011))). See also *Reichle v. Howards*, 566 U.S. 658, 664 (2012) (quoting language to same effect).

⁸⁰ *White*, 137 S. Ct. at 551 (quoting *Mullenix*, 136 S. Ct. at 308 (quoting *Malley v. Briggs*, 475 U.S. 335, 341 (1986))). See also *Sheehan*, 135 at 1774 (quoting same language from *al-Kidd*, 563 U.S. at 743).

⁸¹ *Sheehan*, 135 S. Ct. at 1774.

The Court’s stringency stems from the fundamental purpose of qualified immunity: minimizing the social costs that arise when litigating § 1983 suits. Those costs include deterring competent individuals from serving in government and chilling vigorous decision-making, as well as having to invest time, energy, and money into litigation.⁸² To avoid those costs, qualified immunity doctrine seeks to dispose of insubstantial lawsuits at the summary judgment.⁸³

But while the Supreme Court has insisted that “clearly established” statutory or constitutional rights be defined at a high level of specificity, it has consistently avoided specifying those rights since it overruled *Saucier v. Katz* in *Pearson v. Callahan*. While *Saucier* had required federal courts to determine whether a constitutional violation had been alleged before reaching the question of whether the law was clearly established,⁸⁴ *Pearson* gave them “sound discretion [to] decid[e]” the underlying merits or skip ahead to the qualified-immunity analysis of whether the right was clearly established at the time of the challenged conduct.⁸⁵ Moreover, the Court has restricted the sources from which clearly established law can be drawn.⁸⁶

This has led to a catch-22. In order to hold officers accountable, the law has to be extremely specific. But the Court has balked at creating specific laws because it is often easier

⁸² See, e.g., *Pearson v. Callahan*, 555 U.S. 223, 231, 237 (2009) (quoting *Mitchel v. Forsyth*, 472 U.S. 511, 526 (1985)); *Saucier v. Katz*, 533 U.S. 194, 200 (2001); *Harlow v. Fitzgerald*, 457 U.S. 800, 817 (1982).

⁸³ *Pearson*, 555 U.S. at 231-32 (quoting *Anderson v. Creighton*, 483 U.S. 635, 640 n.2 (1987) and *Hunter v. Bryant*, 502 U.S. 224, 227 (1991)). See also *Harlow*, 457 U.S. at 818; *Saucier*, 533 U.S. at 201. For an important discussion of how qualified immunity fails these policy goals, see Joanna Schwartz, *How Qualified Immunity Fails*, 127 YALE L.J. 2 (2017).

⁸⁴ 533 U.S. at 201.

⁸⁵ 555 U.S. at 236.

⁸⁶ See, e.g., *Carroll v. Carman*, 574 U.S. 13, 17 (2014) (leaving undecided whether a single “controlling circuit precedent could constitute clearly established federal law.”); *Lane v. Franks*, 573 U.S. 228, 246 (2014) (rejecting caselaw on point from other circuits if it conflicts with precedent in the circuit in which plaintiff sued); *al-Kidd*, 563 U.S. at 742 (rejecting that “clearly established law could “lurk[] in the broad history and purposes of the Fourth Amendment” (internal quotations omitted)). See also Kinports, *supra* note 76.

to resolve a case on the demanding clearly-established prong.⁸⁷ This has left “standards of official conduct permanently in limbo,” and given officers considerable leeway to “persist[] in . . . challenged practice[s].”⁸⁸

2. The Current State of Qualified Immunity Scholarship

Qualified immunity has created a slew of interest among academics, journalists,⁸⁹ and interest groups across the political spectrum. Many scholars have criticized qualified immunity on doctrinal or policy grounds. Among the doctrinal critics, William Baude has recently questioned qualified immunity’s legal basis, arguing that Congress never authorized qualified immunity, that its roots in the common law are suspect, that it has become unmoored from *any* common-law standard it once purported to embody, and that it does not withstand scrutiny as a civil equivalent to criminal law’s rule of lenity.⁹⁰ Alan Chen has similarly objected that qualified immunity violates § 1983’s original intent.⁹¹ He has also criticized the Court for “embedding a central paradox” of framing qualified immunity as a pure question of law, while

⁸⁷ Karen M. Blum, *Qualified Immunity: Time to Change the Message*, 93 NOTRE DAME L. REV. 1887, 1905 (2018). For specifics, see *infra*, Sections III.B.ii and III.C. See also Baude, *supra* note 70, at 84-86 (discussing the Court’s extensive use of summary reversals based on the clearly-established prong).

⁸⁸ *Camreta v. Greene*, 563 U.S. 692, 706 (2011) (citation omitted).

⁸⁹ For recent discussions, both before and after the extrajudicial killing of George Floyd, see Eric Schnurer, *Congress Is Going to Have to Repeal Qualified Immunity*, THE ATLANTIC (June 17, 2020), <https://www.theatlantic.com/ideas/archive/2020/06/congress-going-have-repeal-qualified-immunity/613123>; Andrew Chung, et al., *For Cops Who Kill, Special Supreme Court Protection*, REUTERS INVESTIGATES (May 8, 2020, noon, GMT), <https://www.reuters.com/investigates/special-report/usa-police-immunity-scotus>; Mark Joseph Stern, *The Supreme Court Broke Qualified Immunity. Now It Has the Chance to Fix It*, SLATE (May 27, 2020, 5:54PM), <https://slate.com/news-and-politics/2020/05/george-floyd-supreme-court-police-qualified-immunity.html>.

⁹⁰ Baude, *supra* note 70.

⁹¹ Alan K. Chen, *The Facts About Qualified Immunity*, 55 EMORY L.J. 229, 273 (2006).

ignoring the inescapably fact-bound nature of its application.⁹² Kit Kinports has focused on the Court's recent "pattern of covertly broadening [qualified immunity], describing it in increasingly generous terms, and inexplicably adding qualifiers to precedent that then take on a life of their own."⁹³ And Katherine Mims Crocker has analyzed and rejected various justifications for qualified immunity that are grounded in principles separation-of-powers and federalism.⁹⁴

Scholars have also criticized qualified immunity on policy grounds. Joanna Schwartz's empirical work has shown that qualified immunity falls short of its own goals. Because police officers are widely indemnified, qualified immunity does not temper overdeterrence.⁹⁵ It fails to both winnow out insubstantial suits at the motion-to-dismiss and summary-judgment stage⁹⁶ and incentivize plaintiffs' lawyers to decline insubstantial suits.⁹⁷ Instead, it increases the risk, costs, and complexity of constitutional litigation.⁹⁸ In so doing, it fails to protect officers from the financial burdens and distractions of litigation.⁹⁹

⁹² Chen, *The Facts*, *supra* note 91, at 230; *see also* Alan K. Chen, *The Intractability of Qualified Immunity*, 93 NOTRE DAME L. REV. 1937 (2018) (cashing out the implications flowing from that paradox). *Accord*, Blum, *Change the Message*, *supra* note 87, at 1917-32.

⁹³ Kinports, *supra* note 76, at 64. For an earlier critique of the Court's activism in qualified immunity, *see* David Rudovsky, *The Qualified Immunity Doctrine in the Supreme Court: Judicial Activism and the Restriction of Constitutional Rights*, 138 U. PA. L. REV. 23 (1989).

⁹⁴ Katherine Mims Crocker, *Qualified Immunity and Constitutional Structure*, 117 MICH. L. REV. 1405 (2019).

⁹⁵ Joanna C. Schwartz, *Police Indemnification*, 89 N.Y.U. L. REV. 885 (2014). *See also* Barbara E. Armacost, *Qualified Immunity: Ignorance Excused*, 51 VAND. L. REV. 583 (1998).

⁹⁶ Schwartz, *How Qualified Immunity Fails*, *supra* note 83.

⁹⁷ Joanna C. Schwartz, *Qualified Immunity's Selection Effects*, 114 NW. L. REV. 1101 (2020).

⁹⁸ Schwartz, *Selection Effects*, *supra* note 98.

⁹⁹ Joanna C. Schwartz, *The Case Against Qualified Immunity*, 93 NOTRE DAME L. REV. 1797, 1803-14 (2018).

Beyond pointing to qualified immunity's failures on its own terms, critics have also argued that the doctrine stifles the development of constitutional rights,¹⁰⁰ incentivizes government officials to violate rights with impunity,¹⁰¹ and is part of a broader rollback of constitutional rights by the Supreme Court.¹⁰² Based on these powerful critiques, some scholars have called for the abolition of the doctrine¹⁰³ and begun to envision post-qualified-immunity legal regimes.¹⁰⁴

Still, some scholars have argued in support of qualified immunity. Notwithstanding critics' claims to the contrary, John Jeffries maintains that underenforcing money damages claims for constitutional violations encourages courts to develop constitutional law or prevent the rollback of rights.¹⁰⁵ Richard Fallon, too, recently insisted that "the *Harlow* formula . . . is basically sound,"¹⁰⁶ pointing out that abolishing qualified immunity would have unforeseen

¹⁰⁰ Blum, *Change the Message*, *supra* note 87, at 1893-1905. Stephen R. Reinhardt, *The Demise of Habeas Corpus and the Rise of Qualified Immunity: The Court's Ever Increasing Limitations on the Development and Enforcement of Constitutional Rights and Some Particularly Unfortunate Consequences*, 113 MICH. L. REV. 1219, 1244-50 (2015). Schwartz, *Against Qualified Immunity*, *supra* note 99, at 1814-20. *But see* Colin Rolfs, *Qualified Immunity After Pearson v. Callahan*, 59 UCLA L. REV. 468 (2011) (presenting evidence that district courts after *Pearson* have continued to make constitutional rulings in qualified immunity cases).

¹⁰¹ Armacost, *supra* note 95; Reinhardt, *supra* note 100, at 1244-53. David M. Shapiro & Charles Hogle, *The Horror Chamber: Unqualified Impunity in Prison*, 93 NOTRE DAME L. REV. 2021, 2024-35 (2018).

¹⁰² JAMES E. PFANDER, CONSTITUTIONAL TORTS AND THE WAR ON TERROR (2017); Reinhardt, *supra* note 100, at 1245; Nancy Leong, *Making Rights*, 91 B.U. L. REV. 405 (2012); Nancy Leong, *Improving Rights*, 100 VA. L. REV. 377 (2014); Pamela Karlan, *Shoe-Horning, Shell Games, and Enforcing Constitutional Rights in the Twenty-First Century*, 78 UKMC L. REV. 875, 882-88 (2010); Rudovsky, *supra* note 93.

¹⁰³ Blum, *Change the Message*, *supra* note 87, at 1892, 1932-36 (proposing reviving municipal liability under a respondeat superior theory, rendering § 1983 suits against officers in their individual capacity irrelevant and bypassing qualified immunity issues); Schwartz, *Against Qualified Immunity*, *supra* note 99, at 1800.

¹⁰⁴ Joanna C. Schwartz, *After Qualified Immunity*, 120 COLUM L. REV. 309 (2020).

¹⁰⁵ John C. Jeffries, *The Right-Remedy Gap in Constitutional Law*, 109 YALE L.J. 87 (1999) (arguing that underenforcement allows courts to innovate and spell out the law); *see also* Jeffries, *Liability Rule*, *supra* note 70, at 246-49 (renewing the claim). *But see supra* notes 99-101 and accompanying text.

¹⁰⁶ Fallon, *Bidding Farewell*, *supra* note 70, at 989.

ripple effects on substantive rights, standing requirements, and rules of pleading and proof.¹⁰⁷

Aaron Nielson and Christopher Walker have cautioned against getting rid of the defense by emphasizing the importance of statutory stare decisis¹⁰⁸ and federalism.¹⁰⁹

Those defensive accounts tend to be more contextual than qualified immunity's critiques on doctrinal and policy grounds, but they stop short of a crucial question: Why has the Supreme Court chosen to retrench the doctrine time and time again, in spite of mounting criticism and the doctrine's apparent failures? What are its institutional interests in holding on to qualified immunity?

3. The Contribution: An Institutional History and the *Marburian* Move

To understand how qualified immunity ended up in its current dysfunction, and to find a way forward, it is necessary to look at the doctrine's history. This chapter provides that missing history by looking at the doctrine's origins, the institutional pressures and judicial philosophies that shaped its development, and the unforeseen events that precipitated its demise. To focus the analysis, this chapter derives and applies an analytic framework that I term the *Marburian* move. In *Marbury v. Madison*,¹¹⁰ the Court heard an individual's claim against a government officer and concluded that the officer had violated the individual's right. But the Court refused to award a remedy, holding that Congress's grant of original jurisdiction to hear

¹⁰⁷ Richard H. Fallon, Jr., *The Linkage Between Justiciability and Remedies—and Their Connections to Substantive Rights*, 92 VA. L. REV. 633, 639 (2006); Richard H. Fallon, Jr., *Asking the Right Questions About Officer Immunity*, 80 FORDHAM L. REV. 479, 480 (2011). Fallon, who penned the *Harlow* opinions as Justice Powell's clerk, *see infra* note 137 and accompanying text, most recently reaffirmed his position that “the *Harlow* formula . . . is basically sound.” Fallon, *Bidding Farewell*, *supra* note 70, at 961-82.

¹⁰⁸ Aaron L. Nielson & Christopher J. Walker, *A Qualified Defense of Qualified Immunity*, 93 NOTRE DAME L. REV. 1853 (2018).

¹⁰⁹ Aaron L. Nielson & Christopher J. Walker, *Qualified Immunity and Federalism*, 108 GEO. L.J. 229 (2020).

¹¹⁰ 5 U.S. (1 Cranch) 137 (1803).

the case was unconstitutional.¹¹¹ By stressing the importance of individual rights while “rationing” the remedies,¹¹² the Court managed to extricate itself from a contentious political situation. Refusing the power that Congress had granted, the Court created a *new* power—judicial review—that established the independence of the judicial branch.¹¹³

This is what I call the *Marburian* move—a strategy of splitting the difference between individual rights and government accountability, while removing itself from the political fray and building its own judicial power, independence, and legitimacy. I argue that qualified immunity was born out of a *Marburian* move, but that, over time, the doctrine departed from *Marbury*’s wisdom. This *Marburian* lens will allow us to understand the history of qualified immunity from an institutional perspective. Through incremental doctrinal innovations that reacted to inside and outside pressures, the Court fashioned qualified immunity in order to avoid political crossfire and enhance its judicial power, independence, and legitimacy. But over time, the Court foreclosed those doctrinal possibilities, maneuvered itself *into* the political crossfire, and began to jeopardize the independence and legitimacy it had created for itself.

Analyzing qualified immunity from an institutional perspective will serve several purposes. It will explain how qualified immunity ended up in its current, puzzling state and why academic critiques have gone unheard. It will point to a litigation strategy for reforming qualified immunity. Finally, it will offer some insights on the nature of judicial power and the judiciary’s role in our system of government.

¹¹¹ *Id.* at 174-80.

¹¹² Aziz Z. Huq, *Judicial Independence and the Rationing of Constitutional Remedies*, 65 DUKE L.J. 1 (2015). *See also* Akhil Reed Amar, *Of Sovereignty and Federalism*, 96 YALE L.J. 1425, 1515 n.352 (1987) (discussing a “band of underenforcement”).

¹¹³ ROBERT G. MCCLOSKEY, *THE AMERICAN SUPREME COURT* 27-28 (6th ed., 2016).

This chapter proceeds in five Parts. Part I discusses the doctrinal and historical background of *Pierson v. Ray*,¹¹⁴ the origin case of qualified immunity. The discussion will hash out the close parallels between *Pierson* and *Marbury* and develop the analytical framework of the *Marburian* move. Part II traces how the modern qualified immunity standard emerged from *Scheuer v. Rhodes*¹¹⁵ to *Harlow v. Fitzgerald*. Under the stewardship of Justice Byron White, two developments took shape: the qualified immunity standard became a purely objective legal inquiry, and it was uniform across all officers who received it. During that period, the doctrine closely followed *Marbury*'s wisdom by striking a balance between individual rights and selective remedies in order to keep the Court above the political fray. Part III follows the Court's struggle to adhere to a *Marburian* balance between *Harlow* and *Saucier v. Katz*, and its slow departure from it. While non-crucial decisions continued to perform *Marburian* moves, Chief Justice Rehnquist began to undermine the *Marburian* ledger on the rights side. At the same time, Justice Scalia (and others) began to chip away on the remedies side by introducing a particularity requirement for "clearly established" that would become more and more demanding as time went on. *Saucier* was the last gasp in holding the *Marburian* middle ground. But, as Part IV of this chapter describes, the devolution of qualified immunity into indecision and sanctioning of government misconduct became nearly inevitable after the changes that Chief Justice Rehnquist and Justice Scalia's had injected into the doctrine. Part V concludes with a litigation strategy, an explanation for why the Court has ignored academic critiques of qualified immunity, and some reflections on how a *Marburian* analysis informs our understanding of the nature of judicial power and the role of the judiciary.

¹¹⁴ 386 U.S. 547 (1967).

¹¹⁵ 416 U.S. 232 (1974).

I. *PIERSON V. RAY* AND THE MARBURIAN MOVE

1. *Pierson's* Oddly Conservative Holding

The original case of the Supreme Court's long and winding qualified-immunity jurisprudence might have never come before it had it not been for some peculiar circumstances. Petitioner Robert Laughlin Pierson, an Episcopal minister and son-in-law to then-New York State Governor Nelson A. Rockefeller, had been on a "prayer pilgrimage" in the Deep South in September of 1961.¹¹⁶ At the height of the Freedom Rider Movement, Pierson and twenty-seven other priests who were part of the Episcopal Society for Cultural and Racial Unity, were expressing their support for the Freedom Riders by traveling as an integrated group of black and white ministers from New Orleans to Detroit.¹¹⁷ On their way through Jackson, Mississippi, fifteen of them (three Black, twelve white) decided to sit in the "Whites-Only" waiting room to break Mississippi's segregation laws, be arrested, and serve as "witness[es to] focus attention on sham justice in support of segregation."¹¹⁸ When they arrived at the waiting room, local police captain J.L. Ray and two officers promptly arrested them under § 2087.5.1 of the Mississippi Code for congregating "with intent provoke a breach of the peace."¹¹⁹ Local judge Jim Spencer, an Episcopalian himself, handed down the maximum sentence of \$200 and four months in jail, along with a rebuke that it was "the duty of all men who are professors of the Gospel to pay respectful obedience to the civil

¹¹⁶ RAYMOND ARSENAULT, *FREEDOM RIDERS: 1961 AND THE STRUGGLE FOR RACIAL JUSTICE* 433 (2006).

¹¹⁷ ARSENAULT, *supra* note 116, at 433.

¹¹⁸ Reply Brief for Petitioner at 4, *Pierson v. Ray*, 386 U.S. 547 (1967) (Nos. 79, 94), 1966 WL 100721, at *4.

¹¹⁹ MISS. 1942 CODE § 2087.5.1(1), *invalidated by* *Thomas v. Mississippi*, 380 U.S. 524 (1965).

authority.”¹²⁰ Following the dispatch of a now infamous telegram to Attorney General Robert Kennedy and a wave of outrage across the faithful community,¹²¹ the convictions were overturned on appeal for lack of evidence.¹²²

This could have been the end of the matter. Most participants in the broader Freedom Rider movement lacked the funds to affirmatively litigate their rights in court.¹²³ But Robert Pierson and his co-petitioners were different. Alleging wrongful arrest and false imprisonment in violation of their First Amendment Rights to free expression, they filed a damages suit against Judge Spencer, captain Ray, and the arresting officers under 42 U.S.C. § 1983. After a jury found for the defendants, the Fifth Circuit held that the Supreme Court’s recent holding in *Monroe v. Pape*¹²⁴ precluded any executive officer from claiming immunity.¹²⁵ And yet, it also held that no cause of action could be generated from § 1983 in Pierson’s case because the priests’ actions indicated that they had consented to the arrest.¹²⁶ Pierson appealed to the Supreme Court, which reached the same result, but on different grounds. Correcting the Fifth Circuit’s reading of *Monroe*, Chief Justice Warren held for a unanimous Court that § 1983 did

¹²⁰ ARSENAULT, *supra* note 116, at 434.

¹²¹ *Id.* at 433-34.

¹²² Brief for Petitioner at 5, *Pierson v. Ray*, 386 U.S. 547 (1967) (Nos. 79, 94), 1966 WL 100720, at *5 [hereinafter *Pierson Petitioner’s Brief*].

¹²³ ARSENAULT, *supra* note 116, at 439.

¹²⁴ 365 U.S. 167, 184 (1961) (holding that § 1983’s “under color of [state law]” provision reaches conduct by state police officer’s “[m]isuse of power, possessed by virtue of state law and made possible only because the wrongdoer is clothed with the authority of state law” (citation omitted)).

¹²⁵ *Pierson v. Ray*, 352 F.2d 213 (5th Cir. 1965).

¹²⁶ *Id.*

not abrogate a police officer's common-law defense of good faith and probable cause to the common-law claim of false arrest and imprisonment.¹²⁷

Why would the Warren Court, a champion of civil rights and restraining local law enforcement practices,¹²⁸ near-unanimously reverse an appeals court that applied *Monroe* to vindicate the Civil Rights Act of 1871, a statute that had expressly been intended to curb lawless behavior by local judges and law enforcement in the Reconstruction South?¹²⁹ Why recognize a partial immunity for police officers in a case in which a Mississippi state court had dismissed the charges against civil rights activists,¹³⁰ where the segregationist state law in question had been found unconstitutional by the time *Pierson* reached the Court,¹³¹ and where there was evidence that local Jackson police had frequently arrested civil rights activists in efforts to enforce segregation?¹³² The reasons will become clear once we expand the scope of our analysis to include the intricacies of the *Pierson* opinion, the drift of the caselaw at the time, the Warren Court's broader agenda, its institutional constraints, and the political pressures the Court faced at the time.

¹²⁷ *Pierson v. Ray*, 386 U.S. 547, 555 (1967). The opinion was 8-1, but even the lone dissenter, Justice Douglas, only took issue with the separate holding on judicial immunity. *Id.* at 558-67.

¹²⁸ See generally, MORTON HORWITZ, *THE WARREN COURT AND THE PURSUIT OF JUSTICE* (1998) (praising the Warren Court's progressivism); ED CRAY, *CHIEF JUSTICE: A BIOGRAPHY OF EARL WARREN* 289-300, 379-411, 432-42, 455-68 (1997).

¹²⁹ See Alfred Avins, *The Ku Klux Klan Act of 1871: Some Reflected Light on State Action and the Fourteenth Amendment*, 11 ST. LOUIS L.J. 331 (1967) (laying out the Congressional debate). For a compelling *judicial* account of § 1983's purpose to combat lawless conduct by the Ku Klux Klan, shielded by local judges, see *Briscoe v. LaHue*, 460 U.S. 325, 356-64 (1983) (Marshall, J., dissenting). By the time of the Freedom Rides, the Klan's influence in the South was alive and well. ARSENAULT, *supra* note 116, at 425.

¹³⁰ *Pierson Petitioner's Brief*, *supra* note 122, at *5.

¹³¹ *Thomas v. Mississippi*, 360 U.S. 524 (1965).

¹³² ARSENAULT, *supra* note 116, at 433; *Pierson Petitioner's Brief*, *supra* note 122, at *9.

2. *Pierson's* Historical Context

When the Supreme Court handed down *Pierson v. Ray* in 1967, it did not consider the question of officer liability under § 1983 in isolation. Rather, it first considered whether judges' absolute common-law immunity from damages for their judicial acts had survived § 1983, and then considered whether law enforcement officer's common-law defenses in a common-law tort for false arrest and imprisonment had survived § 1983.¹³³ The first question was quickly disposed of: judicial immunity had been a bedrock principle since *Bradley v. Fisher*,¹³⁴ decided in the immediate aftermath of the 1871 Civil Rights Act. If there was any doubt whether § 1983 had intended to abolish that immunity, the Court dispelled it by drawing on its broad holding in *Tenney v. Brandhove*.¹³⁵ The 1871 Congress had not meant to abolish *any* common-law immunities.¹³⁶ The question of police officer immunity thus seemed to answer itself: whichever defense or immunity existed for false arrest at common law would still be available. Chief Justice Warren found that, "[u]nder the prevailing view in this country," that defense consisted in a showing of probable cause that a law had been violated and good faith in its validity.¹³⁷ *Monroe v. Pape* had implied nothing to the contrary, and the defense was available even if the law on which the officer relied for his challenged actions would later be found unconstitutional.¹³⁸ In essence, *Pierson* instituted the principle "[a] policeman's lot is not so

¹³³ The Court also considered the subsidiary questions whether such a defense is available when the state law in question was later found unconstitutional and whether the officers could invoke *another* common-law defense: petitioners' (alleged) consent to arrest. 386 U.S. at 547.

¹³⁴ 80 U.S. 335 (1871).

¹³⁵ 341 U.S. 367 (1951).

¹³⁶ 386 U.S. at 554 (citing *Tenney*). *But see* Avins, *supra* note 129.

¹³⁷ 386 U.S. at 555.

¹³⁸ *Id.* at 555-56.

unhappy that he must choose between being charged with dereliction of duty if he does not arrest when he has probable cause, and being mulcted in damages if he does,”¹³⁹ a proposition that carried much more force at a time when municipal liability was not yet available to plaintiffs.¹⁴⁰

The Supreme Court was not alone in its view. At the time of the *Pierson* decision, five circuit courts had already held that police officers sued under § 1983 for false arrest can invoke the common-law defense of good faith and probable cause.¹⁴¹ The only circuit that had contrary caselaw on the books had overruled itself in light of *Tenney v. Brandhove* and extensive federal caselaw in the lower courts.¹⁴² Another circuit had granted police officers the good-faith and probable-cause defense in § 1983 suits for unconstitutionally misapplying a traffic

¹³⁹ *Id.* at 555.

¹⁴⁰ The Court created municipal liability under § 1983 in 1978. *Monell v. Dep’t of Soc. Servs.*, 436 U.S. 658 (1978). But see James E. Pfander & Jonathan L. Hunt, *Public Wrongs and Private Bills: Indemnification and Government Accountability in the Early Republic*, 85 N.Y.U. L. REV. 1862 (2010) (discussing widespread practice in the early Republic of indemnifying government officials through private bills); Schwartz, *Police Indemnification*, *supra* note 95 (discussing widespread indemnification today).

¹⁴¹ *Joyce v. Ferrazzi*, 323 F.2d 931 (1st Cir. 1963) (finding no § 1983 liability because “not every police error of law or fact arises to the dignity of a deprivation of a federally secured right, privilege, or immunity”); *Gregoire v. Biddle*, 177 F.2d 579, 581 (2d Cir. 1949) (Hand, J.) (“There must indeed be means of punishing public officers who have been truant to their duties; but that is quite another matter from exposing such as have been honestly mistaken to suit by anyone who has suffered from their errors.”); *Mueller v. Powell*, 203 F.2d 797, 800-01 (8th Cir. 1953) (holding officer liable, but stating that, to escape § 1983 liability, “the officer must not act arbitrarily, but must exercise his discretion in a legal manner, using all reasonable means to prevent mistakes. In other words, he must [act] in good faith”); *Pritchard v. Downey*, 326 F.2d 323, 325 (8th Cir. 1964) (exempting officers from § 1983 liability because probable cause for arrest existed), *Agnew v. City of Compton*, 239 F.2d 226, 231 (9th Cir. 1957) (holding that no cause for false arrest under § 1983 was stated, but opining that, in any event, “a law officer’s honest misunderstanding of the law or facts in making an arrest” would not be sufficient); *Beauregard v. Wingard*, 362 F.2d 901, 903 (9th Cir. 1966) (holding that “where probable cause . . . exist[s] civil rights are not violated by an arrest even though innocence may later be established”); *Marland v. Heyse*, 315 F.2d 312, 314 (10th Cir. 1963) (holding that, in a § 1983 suit, it was for the jury to decide whether an arresting officer “was so arbitrary, unreasonable and without probable cause as to subject the plaintiff to deprivation of rights guaranteed by the Constitution of the United States”).

¹⁴² *Bauers v. Heisel*, 361 F.2d 581, 585-88 (3d Cir. 1966) (holding, in light of *Tenney* and vast inter-circuit consensus, that § 1983 was not intended to derogate common-law immunities) (overruling *Picking v. Pa. R. Co.*, 151 F.2d 240 (3d Cir. 1945) (denying immunity under § 1983)).

statute,¹⁴³ while yet another had ruled that police officers could not be sued under § 1983 at all.¹⁴⁴ The Fifth Circuit in *Pierson* was thus an outlier when it ruled that *Monroe v. Pape* foreclosed immunities for police officers sued under § 1983.¹⁴⁵ *Pierson* made a passing reference to the broad circuit consensus, which suggests that it think of its decision as a major innovation.¹⁴⁶ Qualified immunity was thus not a surprising doctrinal aberration. Rather, it emerged as a blessing of a stable inter-circuit consensus that seemed all but inevitable in light of *Bradley v. Fisher* and *Tenney v. Brandhove*.¹⁴⁷

Of course, a discussion of the Supreme Court's modern ratification of police officer immunity cannot be complete without touching on the political backdrop of early 1967. Although the *Pierson* opinion hardly mentions the civil rights context in which the dispute arose, the Court was keenly aware that a suit in which private individuals sought to hold state law enforcement accountable was not purely a legal issue.¹⁴⁸ Mark Tushnet has commented on

¹⁴³ *Hurlburt v. Graham*, 323 F.2d 723, 725 (6th Cir. 1963) (exempting from § 1983 liability police officers who allegedly gave inaccurate testimony in a traffic accident case that applied the wrong statute); *Gabbard v. Rose*, 359 F.2d 182 (6th Cir. 1966) (similar, reasoning that “[n]o one has a constitutional right to be free from a law officer’s honest misunderstanding of law or facts in making an arrest” (quoting *Agnew v. City of Compton*, 239 F.2d 226 (9th Cir. 1956)). *See also* *Downie v. Powers*, 193 F.2d 760 (10th Cir. 1951) (holding that “[d]iligent and conscientious effort is all that is required” of law enforcement officials tasked with keeping the peace).

¹⁴⁴ *Smith v. Dougherty*, 286 F.2d 777 (7th Cir. 1961).

¹⁴⁵ William Baude has critiqued qualified immunity for not being grounded in the common law. Baude, *supra* note 70, at 50-61. While this may be correct, it must be noted that the federal courts at the time of *Pierson* thought otherwise.

¹⁴⁶ 386 U.S. at 555, n.9.

¹⁴⁷ 386 U.S. at 555. *See also* Sheldon Nahmod, *Section 1983 Is Born: The Interlocking Supreme Court Stories of Tenney and Monroe*, 17 LEWIS & CLARK L. REV. 1019 (2013) (discussing Justices Frankfurter and Douglas’s disagreement over the meaning and scope of § 1983 and their varying concern for federalism and individual rights, respectively).

¹⁴⁸ *See, e.g.*, Brief for Respondent at 5, *Pierson v. Ray*, 386 U.S. 547 (1967) (Nos. 79, 94), 1966 WL 115420, at *5 [hereinafter *Pierson Respondent’s Brief*]; *Oral Argument: Pierson v. Ray*, OYEZ (Jan. 11, 1967), at 56:10, <https://www.oyez.org/cases/1966/79>. (“I’d like to get to the immunity of the police officers, because if there is anyone that I feel here . . . that needs representation . . . it is the policeman in these troubled times.”) [hereinafter *Pierson, Respondent’s Brief*].

how “the Warren Court [was] an actor in 1960s politics,”¹⁴⁹ and it certainly felt the political and social pressure that had been mounting ever since its civil rights decisions in *Brown v. Board of Education*,¹⁵⁰ *Heart of Atlanta Motel v. United States*,¹⁵¹ and *Katzbach v. McClung*,¹⁵² as well as its criminal procedure holdings in *Massiah v. United States*,¹⁵³ *Escobedo v. Illinois*,¹⁵⁴ and *Miranda v. Arizona*.¹⁵⁵ *Pierson* was argued on January 11, 1967,¹⁵⁶ just two months after the Democratic party and Lyndon Johnson’s Great Society coalition had suffered a stinging loss in the 1966 midterm elections,¹⁵⁷ despite its recent shift from the war on poverty to the war on crime.¹⁵⁸ “[A]s the Great Society coalition decayed, so did the coherence of the Warren Court,”¹⁵⁹ which Justin Driver convincingly uncovered in his discussion of the Court’s surprisingly conservative

¹⁴⁹ Mark Tushnet, *The Warren Court as History: An Interpretation*, in *THE WARREN COURT IN HISTORICAL AND POLITICAL PERSPECTIVE 2* (ed. Mark Tushnet, 1993); see also L.A. SCOT POWE, *THE WARREN COURT AND AMERICAN POLITICS* (2000).

¹⁵⁰ 347 U.S. 483 (1954) (holding school segregation under the “separate but equal” doctrine unconstitutional).

¹⁵¹ 379 U.S. 241 (1964) (upholding the constitutionality of Title II of the 1964 Civil Rights under the Commerce Clause).

¹⁵² 379 U.S. 294 (1964) (prohibiting that the Commerce Clause empowers Congress to prohibit racial discrimination in restaurants).

¹⁵³ 377 U.S. 201 (1964) (prohibiting the police from deliberately eliciting testimonial statements from a criminal defendant after the right to counsel attaches).

¹⁵⁴ 378 U.S. 478 (1964) (overturning the conviction of a criminal defendant who had been interrogated without counsel in spite of his repeated requests to see his lawyer).

¹⁵⁵ 384 U.S. 436 (1966) (holding that the Fifth Amendment required warnings vindicating the right against self-incrimination before the beginning of a custodial interrogation). For a discussion of public outcries to those cases, see CRAY, *supra* note 128, at 458-60, and POWE, *supra* note 149, at 398-400.

¹⁵⁶ 386 U.S. 547 (1967).

¹⁵⁷ Benjamin J. Guthrie, *Statistics of the Congressional Election of November 8, 1966*, U.S. GOV’T PRINTING OFFICE (1967), http://clerk.house.gov/member_info/electionInfo/1966election.pdf. See also POWE, *supra* note 149, at 400.

¹⁵⁸ ELIZABETH HINTON, *FROM THE WAR ON POVERTY TO THE WAR ON CRIME: THE MAKING OF MASS INCARCERATION IN AMERICA* 27-133 (2016).

¹⁵⁹ Tushnet, *supra* note 149, at 19.

opinions throughout Warren’s tenure, even by the standards of the time.¹⁶⁰ As Ed Cray has pointed out, “the Warren Court was not insensitive to the needs of law enforcement, particularly the safety of police on the streets.”¹⁶¹ And the Court had become especially hesitant to impose further restrictions on the police after the public backlash to *Miranda* in 1966. A mere two years later, Justice Douglas thus called out the Court in his dissent to *Terry v. Ohio* that “hydraulic pressures of this time have made the Court enter a new era and give far more rights and powers to the police.”¹⁶² The times were changing, and soon the Warren Court would hear the Nixon campaign’s clarion calls of “law and order” and “Impeach Earl Warren!”¹⁶³

The birth of qualified immunity to suits under § 1983 therefore makes perfect sense in its legal, political, and institutional context. Having expanded individual rights for years, the Warren Court in *Pierson* eased the pressure on the remedies side by creating a “band of underenforcement” for § 1983.¹⁶⁴ It handed an olive branch to state law enforcement by giving it room to maneuver, while curtailing *Monroe v. Pape* and the reach of a Congressional statute that had given the federal courts vast enforcement powers. And in the midst of a standoff between the people and state governments, the Court had installed itself as an arbiter: Section 1983 would not automatically trigger liability if a constitutional right was violated—the Court had interposed a fault scheme by which courts had room to decide whether officers were

¹⁶⁰ Justin Driver, *The Constitutional Conservatism of the Warren Court*, 100 CALIF. L. REV. 1101 (2012).

¹⁶¹ CRAY, *supra* note 148, at 466.

¹⁶² *Terry v. Ohio*, 392 U.S. 1, 39 (1968) (Douglas, J., dissenting); *see also* POWE, *supra* note 149, at 407.

¹⁶³ *See* POWE, *supra* note 149, at 407 (“*Miranda* was the high point of the Warren Court’s criminal procedure revolution and set the Court on a collision course with the 1968 presidential election.”); *see also* MICHAEL J. GRAETZ & LINDA GREENHOUSE, *THE BURGER COURT AND THE RISE OF THE JUDICIAL RIGHT* 2 (2016); Tushnet, *supra* note 149, at 20.

¹⁶⁴ Amar, *supra* note 112, at 1515 n.352.

shielded. The Court had refrained from exercising its power to impose strict liability—and revived the power to preside over those very same disputes. It is this strategic maneuver that I term the *Marburian* move. A brief analysis of *Marbury v. Madison* will bring this analytical framework into view.

3. *Pierson's Marburian Move*

Robert McCloskey treats *Marbury v. Madison* as paradigm case, “a masterwork of indirection, a brilliant example of . . . advanc[ing] in one direction while [the] opponents are looking in another,” and “of rejecting and assuming power in a single breath.”¹⁶⁵ The facts and holdings of *Marbury* are well-known to every law student.¹⁶⁶ In 1801, William Marbury was denied his commission as justice of the peace after the newly elected President Thomas Jeffersonian instructed his Secretary of State, James Madison, to hold on to all commissions that had not yet been dispatched.¹⁶⁷ Marbury sued for mandamus in the Supreme Court. The Court was facing heavy political pressure at the time, after the outgoing Federalist John Adams had appointed John Marshall as Chief Justice and had filled six newly created appellate courts with sixteen pro-Federalist judges.¹⁶⁸ Chief Justice Marshall, for a unanimous Court, held that, under “a government of laws, and not of men,” Marbury had a legal right to his position, and

¹⁶⁵ MCCLOSKEY, *supra* note 113, at 25, 26.

¹⁶⁶ For more detailed accounts, see WILLIAM E. NELSON, *MARBURY V. MADISON: THE ORIGINS AND LEGACY OF JUDICIAL REVIEW* 54-70 (2000), and WILLIAM H. REHNQUIST, *THE SUPREME COURT* 21-34 (rev'd & updated, 2001). Perhaps the authoritative account of *Marbury* can be found in William W. Van Alstyne, *A Critical Guide to Marbury v. Madison*, 1969 DUKE L.J. 1. On the issue of verifying and pinning down which facts are the important ones, see Sanford Levinson & Jack M. Balkin, *What Are the Facts of Marbury v. Madison?*, 20 CONST. COMMENT. 255 (2003).

¹⁶⁷ Van Alstyne, *supra* note 166, at 4.

¹⁶⁸ *Id.*

that the proper remedy should be mandamus.¹⁶⁹ At the last moment, however the Court withdrew that remedy by holding that the statutory provision authorizing the Court to order mandamus—Section 13 of the 1789 Judiciary Act—was unconstitutional.¹⁷⁰ Handing the executive this victory and backing down from an interbranch confrontation, the Court established the principle of judicial supremacy—designating itself the decider of last resort in disputes over the constitutionality of the nation’s laws.¹⁷¹

Marbury’s genius lay in handing a small victory to an antagonistic executive and “high-mindedly refus[ing]” a statutory grant of jurisdiction from an allied Congress,¹⁷² all the while creating an independent power of its own. The Court did so by insisting on the broad availability of individual rights, but being elusive on the remedy. The *Marburian* move averted a political standoff, created a legal innovation, and strengthened the Court as an institution.

Pierson v. Ray near-perfectly replicates the *Marburian* move. It handed a small victory to law enforcement, declined to exercise the broad statutory grant by the 1871 Civil Rights Act to enforce federal rights against lawless state actors, and gave courts the last word on whether police officers should be held liable for violating an individual’s federal rights.¹⁷³ Here, too, the Court backed down in a time of political pressure by introducing a legal innovation. *Pierson* was a *Marburian* move par excellence—a political gambit, a doctrinal innovation, and an

¹⁶⁹ *Marbury v. Madison*, 5 U.S. (1 Cranch) 137, 163, 173 (1803).

¹⁷⁰ *Id.* at 174-80.

¹⁷¹ *Id.* at 180.

¹⁷² MCCLOSKEY, *supra* note 113, at 26.

¹⁷³ An unbridled application of *Monroe v. Pape* to give effect to § 1983’s strict liability scheme would not have given the judiciary as much room to maneuver as it did under the fault scheme of probable cause and good faith. The difference between *Marbury* and *Pierson* is of course that the Court held a congressional act unconstitutional in the former, and only created a defense to one in the latter. But this is a difference in degree, not in kind, especially given the later developments that progressively hollowed out § 1983. The other—for our purposes, inconsequential—difference is the statutory remedy involved (mandamus versus money damages).

institutional fortification. For the rest of this chapter, the *Marburian* framework will guide our historical analysis of qualified immunity. It will uncover how the Court’s qualified immunity decisions initially followed *Marbury*’s wisdom—and how they came to abandon that wisdom as time went on.

II. FOLLOWING *MARBURY* (MOSTLY): CRAFTING A NEW STANDARD FROM *SCHEUER V. RHODES* TO *HARLOW V. FITZGERALD*

This Part will trace how the Court fashioned the modern qualified-immunity standard. In *Scheuer v. Rhodes*,¹⁷⁴ the Court expanded *Pierson*’s fairly narrow holding into the broad, unspecific proposition that executive officers should receive *some* form of immunity for their official acts. As the docket of federal civil rights cases grew, *Scheuer* raised many questions about the contours and applicability of the new standard. The Court, under the stewardship of Justice Byron White, addressed these questions by shaping the standard in two major ways: making it wholly objective and making it uniform for every official performing executive acts.

1. *Scheuer v. Rhodes*: Broadening the Defense and Raising New Questions

A. An Atmosphere “Rife with Political Passion”

Seven years after *Pierson*, the Supreme Court confronted the question of officer immunity in the civil damages context again in *Scheuer v. Rhodes*. *Scheuer* arose from the circumstances surrounding the infamous Kent State University shooting on May 4, 1970. At the time, the war in Vietnam had taken center stage,¹⁷⁵ and when Nixon announced on April

¹⁷⁴ 416 U.S. 232 (1974).

¹⁷⁵ BRUCE J. SCHULMAN, THE SEVENTIES: THE GREAT SHIFT IN AMERICAN CULTURE, SOCIETY, AND POLITICS 14 (2002).

28 that the United States had expanded the war into Cambodia, student protests broke out across the country.¹⁷⁶ Nixon, in characteristic fashion, referred to those protesters as “bums.”¹⁷⁷ The protests came at an inopportune moment for Ohio Governor Jim A. Rhodes, who had been running as a law-and-order candidate in a hotly contested Senatorial primary race at the time.¹⁷⁸ Rhodes called on the National Guards to quell the Wildcat trucker strike on April 29.¹⁷⁹ When protestors burned down an ROTC house on May 2,¹⁸⁰ the governor publicly vowed to “eradicate the problem”¹⁸¹ and gave Adjutant General Sylvester Del Corso the authority to move the troops onto Kent State’s campus.¹⁸² In the hours that followed, nine students were injured and four were shot dead.¹⁸³

In this atmosphere, “rife with political passion”¹⁸⁴ and marred by “local passions and manipulation,”¹⁸⁵ the courts could not resist siding with the state. When the victims’ decedents

¹⁷⁶ *Id.* at 43-44.

¹⁷⁷ Juan de Onis, *Nixon Puts “Bums” Label on Some College Radicals*, N.Y. TIMES A1 (May 2, 1970), <https://www.nytimes.com/1970/05/02/archives/nixon-puts-bums-label-on-some-college-radicals-nixon-denounces-bums.html>.

¹⁷⁸ Brief for Petitioner at 59, *Scheuer v. Rhodes*, 416 U.S. 232 (1974) (No. 72-1318), 1972 WL 135764, at *59 [hereinafter *Scheuer Petitioner’s Brief*].

¹⁷⁹ *Scheuer Petitioner’s Brief*, *supra* note 178, at *61.

¹⁸⁰ Brief for Amicus Curiae on Behalf of the National Council of Churches of Christ in the U.S.A. et al. at 5, *Scheuer v. Rhodes*, 416 U.S. 232 (1974) (Nos. 72-1318 and 72-914), 1973 WL 171712, at *5 [hereinafter *Scheuer Church Brief*].

¹⁸¹ *Id.* at *6.

¹⁸² *Id.* at *5-6.

¹⁸³ John Kifner, *4 Kent State Students Killed by Troops*, N.Y. TIMES A1 (May 5, 1970),

<https://www.nytimes.com/1970/05/05/archives/4-kent-state-students-killed-by-troops-8-hurt-as-shooting-follows.html>

¹⁸⁴ *Scheuer Petitioner’s Brief*, *supra* note 178, at *59.

¹⁸⁵ Brief for Amicus Curiae Kent Legal Defense Fund at 2, *Scheuer v. Rhodes*, 416 U.S. 232 (1974) (No. 72-914), 1972 WL 135870, at *2.

brought a wrongful death suit under § 1983 against the Governor and various Guardsmen, the district court dismissed the complaints before the defendants had even filed their answers, holding that the suit effectively against the State of Ohio and thus barred by the Eleventh Amendment.¹⁸⁶ In affirming the district court's holding, the Sixth Circuit added that the Governor and Guardsmen were entitled to absolute executive immunity, thereby accepting the lower court's judicial notice that a state of insurrection existed at the time.¹⁸⁷ As government and citizens accused each other of lawlessness, the Supreme Court was called upon as the final arbiter between the state authorities and the people.

B. Scheuer's Marburian Move

The *Scheuer* Court performed another deft *Marburian* move by navigating the case's political and institutional pitfalls. It allowed the victims to plead their cases in court, but left room to maneuver on the remedies side by extending the reach of qualified immunity. As an opening gambit, *Scheuer v. Rhodes* rebuffed the state sovereign immunity argument. Extending *Ex parte Young*¹⁸⁸ from the injunctive context to the damages context, the Court held that the Eleventh Amendment did not bar suits against officials in their individual capacity, reasoning that state officers who violate the Constitution were "stripped of [their] official or representative character and [are] subjected . . . to the consequences of [their] individual

¹⁸⁶ *Scheuer Petitioner's Brief*, *supra* note 178, at *8.

¹⁸⁷ *Krause v. Rhodes*, 471 F.2d 430, 432-34 (6th Cir. 1972). *See also* *Krause Scheuer Petitioner's Brief*, *supra* note 178, at *19.

¹⁸⁸ 209 U.S. 123 (1908).

conduct.”¹⁸⁹ In the next breath, however, Chief Justice Burger, for a unanimous Court,¹⁹⁰ handed an olive branch to state governments. While executive officials’ immunity cannot be absolute, it had to vary with “the scope of discretion and responsibilities of the office and all the circumstances as they reasonably appeared at the time.”¹⁹¹ Courts could consider whether there were “reasonable grounds for the belief formed at the time . . . in light of all the circumstances” and whether the officer had a “good-faith belief”¹⁹² that his conduct was lawful.

This deliberately hazy standard not only managed to be even-handed—the Court had overturned two lower courts that had run away with the politics of the case—, but it also enhanced the power of the courts. Federal courts became the forum in which to defuse and depoliticize contentious clashes between the people and state authorities.¹⁹³ By splitting the difference between the two parties, the Court again curtailed § 1983’s jurisdictional grant, made yet another advance in its new qualified immunity doctrine, increased the judiciary’s power and independence, and avoided the political crossfire. Indeed, the *New York Times* lauded *Scheuer* for “uphold[ing] the [decedents’] right . . . to a trial on the merits of charges that their civil rights had been violated in the campus demonstrations.”¹⁹⁴ Meanwhile the decedent

¹⁸⁹ *Scheuer v. Rhodes*, 416 U.S. 232, 237 (1974) (internal quotation marks omitted).

¹⁹⁰ *Id.* at 232. Justice Douglas had recused himself. Warren Weaver, Jr., *Parents of Kent Victims Can Sue, High Court Says*, N.Y. TIMES A1 (Apr. 18, 1974), <https://www.nytimes.com/1974/04/18/archives/parents-of-kent-victims-can-sue-high-court-says-parents-of-victims.html>.

¹⁹¹ *Id.* at 247.

¹⁹² *Id.* at 247-48.

¹⁹³ Compare the *Scheuer* Court’s cryptic reference to the events as “the same period of alleged civil disorder on the campus of Kent State University in Ohio during May 1970 which was before us, in another context, in *Gilligan v. Morgan*,” *id.* at 234 (citation omitted) with the lower courts’ judicial notice that there was “a mob in insurrection”. *Krause v. Rhodes*, 471 F.2d 430, 435 (6th Cir. 1972).

¹⁹⁴ Weaver, *supra* note 190.

Scheuer's mother was "very pleased that we're finally getting something done."¹⁹⁵ And the defendant Adjutant General Del Corso remarked that he was "not worried, but [also] not elated" by the Court's decision.¹⁹⁶

C. New Questions and a Marburian Mover

From an institutional standoff, the Court had emerged with enhanced judicial power. But now that the courthouse doors had been opened, the Court's new, undefined standard raised a host of new questions. The imprecise content of the immunity standard would force the Court to confront three questions. First, should courts emphasize the objective component of reasonable grounds, or the subjective component of the officer's good-faith intent?¹⁹⁷ Second, how would the immunity standard "vary[] [by] the scope of discretion and responsibilities of the office"?¹⁹⁸ And third, how should courts reliably distinguish executive officials from legislative or judicial officers when their functions overlapped?¹⁹⁹ The Court would also have to grapple with how the immunity standard would fit into the larger legal landscape. The most pressing questions here were: Should the immunity standard respond to the increased load of § 1983 cases that came in the wake of the Court's expansion of constitutional and statutory rights?²⁰⁰ Should qualified immunity take federalism concerns into account by striking a proper balance between federal judicial power and state government

¹⁹⁵ *Id.*

¹⁹⁶ *Id.*

¹⁹⁷ *See infra*, Section II.B.

¹⁹⁸ *See infra*, Section II.C.

¹⁹⁹ *See id.*

²⁰⁰ *See infra*, Section II.B.ii.

actors?²⁰¹ And should the immunity regimes for federal officials sued under *Bivens* and state officials sued under § 1983 be the same or distinct?²⁰²

The Court would find answers to all these questions in the next eight years, when it decided the twin landmark cases *Nixon v. Fitzgerald* and *Harlow v. Fitzgerald*. The main driving force behind that development was Justice Byron White.²⁰³ As the Court’s “extremely influential” center Justice in 1970s and early 1980s, his “effect on the Court’s agenda as well as its direction had never been greater.”²⁰⁴ Keenly aware of the Court’s institutional concerns,²⁰⁵ Justice White spearheaded two major developments in qualified immunity doctrine: turning the standard into a purely objective one, and ensuring that the standard would be uniform across all officials performing executive acts. The cases that advanced those two developments hewed closely to the wisdom of *Marbury*.

2. Towards an Objective Standard: The “Promise of § 1983”

On March 1, 1982, Justice Powell sent a memorandum of proposed changes in *Harlow v. Fitzgerald* to his law clerk Richard Fallon. The memo had a post-it note stuck to it that read: “Dick – Having to make the changes suggested, after your long and thoughtful labor on this

²⁰¹ See *infra*, Section II.C.

²⁰² See *id.*

²⁰³ See *infra*, Sections II.B., II.C., III.A. Justice White’s deep influence seems to have been overlooked in the few discussions that touch on qualified immunity’s history. See, e.g., Crocker, *supra* note 94, at 1429-31 (focusing on “Justice Powell’s significance”).

²⁰⁴ DENNIS J. HUTCHINSON, THE MAN WHO ONCE WAS WHIZZER WHITE: A PORTRAIT OF JUSTICE BYRON R. WHITE 382 (1998).

²⁰⁵ HUTCHINSON, *supra* note 204, at 346, 355, 365, 373-75.

case, is enough to make you resign. Please don't! LFP”²⁰⁶ Powell had been agonizing over *Harlow* and its companion case, *Nixon v. Fitzgerald*, for several months and had been struggling to put together strong coalitions for them. But in early March, Powell saw a way out of the deadlock by adopting an entirely objective qualified-immunity standard, an idea he adopted from Justice Byron White. In his memo to Fallon, Powell wrote: “[A]s [Justice White] moved to this position last Term in his memorandum in [*Halperin*], and seems firmly of this opinion, we probably would have a badly fractured Court if I retain the ‘malice’ component. Byron thinks – and he may be right – that the Chief and possibly even Rehnquist – would join this reformulation of the standard.”²⁰⁷ In hope of a strong Court majority, Powell then set out to woo the other Justices.²⁰⁸

But Justice White’s advocacy for a purely objective standard was not of recent vintage, as Justice Powell suggested to Fallon. In a communication to Powell that copied the whole conference, White remarked on May 21: “You will . . . recall that in early circulations in [*Procunier v. Navarette*], I unsuccessfully proposed eliminating the good faith-malicious intention prong of the *Scheuer-Strickland* formulation of the qualified immunity test. That requirement has never made a great deal of sense to me. At the end of my memorandum last term in

²⁰⁶ Note from Lewis F. Powell, Jr., Assoc. Justice, U.S. Supreme Court, to Richard H. Fallon, Law Clerk, United States Supreme Court (Mar. 1, 1982), in 84 POWELL PAPERS 2, <https://scholarlycommons.law.wlu.edu/cgi/viewcontent.cgi?article=1139&context=casefiles>.

²⁰⁷ Memorandum from Lewis F. Powell, Jr., Assoc. Justice, U.S. Supreme Court, to Richard Fallon, Law Clerk, United States Supreme Court (Mar. 1, 1982), in 84 POWELL PAPERS, *supra* note 206, at 3-4.

²⁰⁸ See, e.g., Letter from Lewis F. Powell, Jr., Assoc. Justice, U.S. Supreme Court, to John Paul Stevens, Assoc. Justice, U.S. Supreme Court (Mar. 15, 1982) (“I consider your support essential. Sandra [Day O’Connor] favors qualified immunity. But, unless the opinion persuades the Chief and Bill Rehnquist, they will go for absolute immunity. Whether Byron will be content with the way I have written *Harlow* remains to be seen. If he joins us, we will still need one of our Brother who were with Byron last Term.”) in 84 POWELL PAPERS, *supra* note 206, at 13; Memorandum from Richard H. Fallon, Law Clerk, U.S. Supreme Court, to Lewis F. Powell, Jr., Assoc. Justice, U.S. Supreme Court (May 27, 1982) (“What to do? [Justice Brennan] has not joined the opinion. According to his clerks, he and several others are awaiting the lead of [Justice White].”), in 84 POWELL PAPERS, *supra* note 206, at 96.

[*Halperin v. Kissinger*,] I also suggested that the rule be modified and have renewed that suggestion to you earlier this term.”²⁰⁹ In fact, Justice White’s push for a purely objective standard had not begun in *Procunier v. Navarette*,²¹⁰ but in *Wood v. Strickland*,²¹¹ one year after *Scheuer*.

A. *Wood v. Strickland*: *Splitting the Prongs*

Wood v. Strickland was “an altogether extraordinary case [in which e]veryone seem[ed] to have cast reason to the winds.”²¹² On February 18, 1972, Peggy Strickland, Virginia Crain, and Jo Wahl, were expelled from their high school in Polk County, Arkansas, for spiking the punch at a school-sponsored function of *Future Homemakers of America*.²¹³ With a bottle of 3.2% malt liquor, the teenagers had created a concoction with an alcohol content of merely 0.91%.²¹⁴ Nonetheless, the schoolboard expelled them for having violated the school regulation prohibiting the use of an “intoxicating beverage” at a school-sponsored function.²¹⁵ When the schoolboard refused to reconsider its decision, Peggy’s and Virginia’s parents filed a § 1983 suit, alleging a deprivation of due process.²¹⁶ The district court directed verdicts for defendants

²⁰⁹ Letter from Byron R. White, Assoc. Justice, U.S. Supreme Court, to Lewis F. Powell, Jr., Assoc. Justice, U.S. Supreme Court (May 21, 1982), in 84 POWELL PAPERS, *supra* note 206, at 77. *See also* Memorandum from Byron R. White, Assoc. Justice, U.S. Supreme Court, to the Conference of the United States Supreme Court on *Kissinger v. Halperin* 35-38 (recirculated Apr. 9, 1981), in 84 POWELL PAPERS, *supra* note 206, at 66-69.

²¹⁰ 434 U.S. 555 (1978).

²¹¹ 420 U.S. 308 (1975).

²¹² Bench Memorandum from Ron Carr, Law Clerk, U.S. Supreme Court, to Lewis F. Powell, Jr., Assoc. Justice, U.S. Supreme Court (Oct. 14, 1974), 414 POWELL PAPERS 15, <https://scholarlycommons.law.wlu.edu/cgi/viewcontent.cgi?article=1710&context=casefiles>.

²¹³ Brief for Respondent at 5-6, *Wood v. Strickland*, 420 U.S. 308 (1975) (No. 73-1285), 1974 WL 186252, at *5-6.

²¹⁴ *Id.* at *5-7.

²¹⁵ *Id.* at *7-12. *See also Wood*, 420 U.S. at 311 & n.3.

²¹⁶ *Strickland v. Inlow*, 384 F. Supp. 244 (W.D. Ark. 1972).

because the parents failed to show malice under the qualified immunity’s good-faith requirement standard.²¹⁷ To the district court, the qualified-immunity standard was entirely subjective.²¹⁸ For the Eighth Circuit, however, “[t]he test [was] an objective, rather than a subjective, one”: the existence of good faith was to be determined by looking at all the circumstances under a general reasonableness standard.²¹⁹ The Eighth Circuit also held that, having misconstrued its own regulation, the schoolboard had deprived the school girls of substantive due process and reversed,²²⁰ sending gasps of disbelief through Justice Powell’s chambers.²²¹

Justice White seized on the confusions and opportunities presented by this case to begin making the qualified-immunity standard more objective. His innovation was to clarify the general *Scheuer* standard by splitting it into two prongs. To be entitled to immunity, a school board official had to show that “he [neither] knew [nor] reasonably [could] have known that [his] action . . . would violate the constitutional rights of the student affected” and that he did not “[take] the action with the malicious intention to cause a deprivation of constitutional rights or other injury to the student.”²²² “Any lesser standard,” Justice White contended, “would deny much of the promise of § 1983.”²²³

²¹⁷ *Id.* at 247, 250-52.

²¹⁸ *Id.* at 250-51.

²¹⁹ *Strickland v. Inlow*, 485 F.2d 186, 191 (8th Cir. 1973).

²²⁰ *Id.* at 190.

²²¹ See Bench Memorandum from Ron Carr, in 414 POWELL PAPERS, *supra* note 212, at 15.

²²² *Wood v. Strickland*, 420 U.S. 308, 322 (1975).

²²³ *Id.*

White's move, however, capitalized on a false dichotomy between the district and appellate courts' decisions. The Eighth Circuit's objective standard had been an outlier among the federal courts.²²⁴ Coming down the middle between the Eighth Circuit and the district court was thus an unprecedented move towards objectivity, as commentators observed.²²⁵ Justice Powell, too, picked up on this innovation and, along with three other Justices, dissented precisely on this point, arguing that it was a new requirement—and an onerous one at that.²²⁶ Did some 20,000 school board officials have to keep abreast of an area of constitutional law that was rapidly developing at the time?²²⁷

B. *Broader Concerns*

At first blush, Justice White's new objective prong may seem unwise from a *Marburian* perspective. While an in-depth discussion of the due process revolution is beyond the scope of this chapter, it was clear to scholars and judges at the time that federal rights had been broadly and innovatively expanded under procedural and substantive due process.²²⁸ Indeed, shortly after *Wood* had been briefed and argued, the Supreme Court recognized education as a protected liberty interest under the Fourteenth Amendment in *Goss v. Lopez*.²²⁹ The Court thus

²²⁴ See *id.* at 315 & n.7 (listing cases that are all wholly or partially subjective).

²²⁵ See, e.g., Patricia L. Stearns, *Wood v. Strickland: Objectifying the Standard of Good Faith for School Board Members in Defense to Personal Liability Under Section 1983*, 10 LOY. L.A. L. REV. 149, 158, 167-70 & nn.72-87 (1976); Richard Briffault, *Section 1983 and Federalism*, 90 HARV. L. REV. 1133, 1212-17 (1977).

²²⁶ 420 U.S. at 329 (Powell, J. dissenting).

²²⁷ *Id.* at 331.

²²⁸ Erwin N. Griswold, *The Due Process Revolution and Confrontation*, 119 U. PA. L. REV. 711 (1971); William G. Buss, *Procedural Due Process for School Discipline: Probing the Constitutional Outline*, 119 U. PA. L. REV. 545 (1971); Henry J. Friendly, "Some Kind of Hearing," 123 U. PA. L. REV. 1267, 1268 (1975); see also JERRY L. MASHAW, *DUE PROCESS IN THE ADMINISTRATIVE STATE* 9 (1985).

²²⁹ 419 U.S. 565, 572-76 (1975).

found itself at an early crossroads in its qualified-immunity jurisprudence: as federal rights expanded and civil rights dockets ballooned, more and more § 1983 cases came into the federal courts.²³⁰ Theodore Eisenberg observed that, in light of expanding federal rights, “there are at least two competing visions of section 1983”: restricting it to “address[] a limited historical problem in post-Civil War race relations” on the one hand, and making it “the primary civil mechanism for vindicating all constitutional rights” on the other.²³¹ Justice White’s addition of the objective prong thus seemed to cut against *Marbury*’s wisdom: as the Court was holding up the banner for federal rights, it seemed to deprive itself of a safety valve in making it harder for defendants to obtain immunity.²³²

Relatedly, some commentators and conservative members of the Court began to raise federalism concerns. A permissive § 1983 liability scheme to enforce broad-sweeping constitutional rights threatened to draw traditionally local decision-making powers—such as school discipline—into the federal forum.²³³ Even more worryingly, if a state or local claim was lost, § 1983 seemed to open an avenue to relitigating those claims in federal court, so long as a constitutional hook could be found.²³⁴ According to Justice Rehnquist, combining § 1983

²³⁰ See Stearns, *supra* note 225, at 149-53 & nn.1-13 (discussing the trend and collecting cases); Theodore Eisenberg, *Section 1983: Doctrinal Foundations and an Empirical Study*, 67 CORNELL L. REV. 482, 483 (1982) (“Rapid expansion of constitutional guarantees inevitably strains a provision that associates a private damages action with each new constitutional right.”); Brief for Petitioner at 16 & n.24, *Procunier v. Navarette*, 434 U.S. 555 (1978) (No. 76-446), 1977 WL 189397, at *16 & n.24 (discussing increase federal civil rights caseloads: approximately 300 in 1960, and 8,400 in 1974).

²³¹ Eisenberg, *supra* note 230, at 483.

²³² *Wood v. Strickland*, 420 U.S. 308, 330-31 (1975) (Powell, J., dissenting).

²³³ Brief for Petitioner at 22-23, *Wood v. Strickland*, 420 U.S. 308 (1975) (No. 73-1285), 1974 WL 186251, at *22-23; see also *Developments in the Law: Academic Freedom*, 81 HARV. L. REV. 1045, 1051-55 (1968).

²³⁴ *Paul v. Davis*, 424 U.S. 693, 699-701 (1976). At oral argument for *Wood*, the slippage of federal courts into the position of de novo review tribunals of local decisions became readily apparent in an exchange among the Justices.

with the due process revolution risked “mak[ing] of the Fourteenth Amendment a font of tort law to be superimposed upon whatever systems may already be administered by the States.”²³⁵ This backdrop, which was by no means restricted to the education context,²³⁶ illustrates the Court’s institutional concerns that were under pressure at the time it decided *Wood v. Strickland*: docket management and diminishing judicial independence; potential politicization of the courts; and threats to judicial legitimacy in becoming reviewing bodies for state judicial decisions. At first blush, Justice White’s addition of an objective component to the qualified-immunity inquiry thus ran counter to *Marbury*’s wisdom, as it seemed to place a thumb on the scale for plaintiffs on the remedies side while *also* giving them wide latitude on the rights side.

C. *Procunier v. Navarette* and *Harlow v. Fitzgerald*: *Securing the Objective Standard*

But the *Marburian* wisdom of Justice White’s move began to become apparent in *Procunier v. Navarette*,²³⁷ the decision in which White initially attempted to drop the subjective

Stewart: It is a question, isn’t it, whether or not the girls engaged in conduct that is proscribed by the rule of the School Board 3(b)?

Mr. Core [Counsel for Respondent]: That’s true.

Stewart: That’s the question.

Rehnquist: Well, or more properly, to what extent should a federal court in a 1983 action second-guess a school board’s decision that that regulation was violated?

Oral Argument: Wood v. Strickland, OYEZ (Oct. 16, 1974), at 34:25, <https://www.oyez.org/cases/1974/73-1285>.

²³⁵ *Paul*, 424 U.S. at 701 (1976).

²³⁶ Stearns, *supra* note 225, at 149-51.

²³⁷ 434 U.S. 555, 561 (1978). Here, a prisoner had filed a § 1983 suit against state prison officials for “negligently or inadvertently interfered with his outgoing mail” and thus violated his First Amendment right to free expression and his due process rights. *Id.* at 557-58.

component.²³⁸ Although he was ultimately unsuccessful, he still managed to deemphasize the subjective prong by holding that claiming mere negligence could never meet the malice requirement.²³⁹ At the same time, he made the objective prong more permissible by lifting *Wood*'s dictum that the allegedly violated constitutional right had to be "clearly established"²⁴⁰ and embedding it into the qualified-immunity standard.²⁴¹ This tightening was another innovative move, as litigation involving the qualified immunity defense tended to focus on whether the allegedly negligent conduct created a § 1983 cause of action that could overcome the subjective immunity prong.²⁴² Justice White continued to thread the needle in *Kissinger v. Halperin*,²⁴³ in which his vigorous and voluminous arguments in favor of an objective standard eroded Justice Powell's bare majority and forced a four-to-four split.²⁴⁴

When Justice Powell decided to drop the subjective prong in the late drafting stages of *Harlow v. Fitzgerald*,²⁴⁵ the change did not happen on a whim. Rather, it was the result of an eight-year effort by Justice White, aided by several contingent factors. First, the Court was just

²³⁸ Letter from Byron R. White, Assoc. Justice, U.S. Supreme Court, to Lewis F. Powell, Jr., Assoc. Justice, U.S. Supreme Court (May 21, 1982), in 84 POWELL PAPERS, *supra* note 206, at 77.

²³⁹ *Navarette*, 434 U.S. at 557-58.

²⁴⁰ *Wood v. Strickland*, 420 U.S. 308, 322 (1975).

²⁴¹ *Navarette*, 434 U.S. at 562. The right now had to be "clearly established at the time of [the] challenged conduct." *Id.*

²⁴² The question that the Supreme Court certified was "[w]hether negligent failure to mail certain of a prisoner's outgoing letters states a cause of action under section 1983?" *Id.* at 570 n.6. *But id.* at 566-56 (Burger, C.J., dissenting) ("I dissent because the Court's opinion departs from our practice of considering only the question upon which certiorari was granted or questions 'comprised therein.'" (citation omitted))

²⁴³ 452 U.S. 713 (1981).

²⁴⁴ GRAETZ & GREENHOUSE, *supra* note 163, at 319; *see generally* Paul J. Wahlbeck et al., *Kissinger v. Halperin*, 452 U.S. 713 (1981), THE BURGER COURT OPINION WRITING DATABASE, <http://supremecourttopinions.wustl.edu/index.php?rt=pdfarchive/details/2091> [hereinafter *Halperin Papers*].

²⁴⁵ The Court had granted certiorari in this case on June 22, 1981, heard oral argument on November 30, 1981 and announced its decision on June 24, 1982.

as splintered on the qualified-immunity standard for high-ranking officials as it had been in *Halperin*. As Justice Powell confided to his law clerk Richard Fallon, “we probably would have a badly fractured Court if I retain the ‘malice’ component.”²⁴⁶ But, persuaded now that White’s position on a purely objective standard was “basically right,” Powell reasoned that “Byron thinks . . . that the Chief and possibly even Rehnquist – would join this reformulation of the standard [in objective terms].”²⁴⁷ An objective standard provided the way out of a deadlock that Powell had been struggling to break for months. The prospect of a strong majority was even more compelling to Powell when considering that, in the companion case *Nixon v. Fitzgerald*, the Justices had fiercely split along ideological lines.²⁴⁸

Second, Justice Powell was especially eager to compromise if he could get Justice Rehnquist on board, who had taken the position four years earlier that all federal executive officials should receive absolute immunity.²⁴⁹ Making the standard an objective one would turn qualified immunity into a purely legal analysis that would facilitate the elimination of insubstantial lawsuits before they went to trial²⁵⁰—a key concern for Justice Rehnquist.²⁵¹ Third, Justice White exerted his influence on the Court to bring Justices Brennan and Marshall

²⁴⁶ Memorandum from Lewis F. Powell, Jr., Assoc. Justice, U.S. Supreme Court, to Richard Fallon, Law Clerk, U.S. Supreme Court (Mar. 1, 1982), in 84 POWELL PAPERS, *supra* note 206, at 3-4.

²⁴⁷ *Id.*

²⁴⁸ See *infra*, Section II.C.ii.

²⁴⁹ *Butz v. Economou*, 438 U.S. 478, 517-18, 525-30 (1978) (Rehnquist, J., dissenting).

²⁵⁰ *Harlow v. Fitzgerald*, 457 U.S. 800, 817 (1982).

²⁵¹ See *supra* note 235 and accompanying text.

into the fold.²⁵² In the end, Justice Powell, with Justice White's support,²⁵³ was able to draw together an eight-member majority to sign on to the new standard that "government officials performing discretionary functions, generally are shielded from liability for civil damages insofar as their conduct does not violate clearly established statutory or constitutional rights of which a reasonable person would have known."²⁵⁴

Justice Powell, driven by "strong institutional reason[s] for avoiding fractionalization,"²⁵⁵ managed to perform a *Marburian* move in *Harlow*: The objective standard, which explicitly sought to balance "the vindication of constitutional rights" against government efficiency,²⁵⁶ was a middle ground between absolute immunity that some members of the Court had favored, and a more permissive qualified-immunity standard. It drew together a strong majority and managed to avoid the charge of being a politically driven outcome.²⁵⁷ *Harlow* also created a new judicial authority: the objective standard turned the immunity question into a question of law that was for judges to decide, not juries.²⁵⁸

²⁵² See Memorandum from Richard H. Fallon, Law Clerk, U.S. Supreme Court, to Lewis F. Powell, Jr., Assoc. Justice, U.S. Supreme Court (May 27, 1982) ("What to do? [Justice Brennan] has not joined the opinion. According to his clerks, he and several others are awaiting the lead of [Justice White]."), in 84 POWELL PAPERS, *supra* note 206, at 96.

²⁵³ See Memorandum from Lewis F. Powell, Jr., Assoc. Justice, U.S. Supreme Court, to Richard Fallon, Law Clerk, U.S. Supreme Court (Mar. 1, 1982), in 84 POWELL PAPERS, *supra* note 206, at 3-4.; see also Letter from Lewis F. Powell, Jr., Assoc. Justice, U.S. Supreme Court, to Byron R. White, Assoc. Justice, U.S. Supreme Court (June 7, 1982), in 84 POWELL PAPERS, *supra* note 206, at 115 ("I am grateful to you for your help. At this season of year, one's own problems more than suffice to overwhelm. At least mine do!").

²⁵⁴ *Harlow*, 457 U.S. at 818.

²⁵⁵ Draft Letter from Lewis F. Powell, Jr., Assoc. Justice, U.S. Supreme Court, to Warren E. Burger, Chief Justice, U.S. Supreme Court (Dec. 16, 1981), in 81 POWELL PAPERS 120, <https://scholarlycommons.law.wlu.edu/cgi/viewcontent.cgi?article=1134&context=casefiles>.

²⁵⁶ *Harlow*, 457 U.S. at 814.

²⁵⁷ See Linda Greenhouse, *High Court Holds President Immune from Damage Suits*, N.Y. TIMES A12 (June 25, 1982), in 81 POWELL PAPERS, *supra* note 255, at 239 (focusing on *Nixon v. Fitzgerald*), and Fred Barbash, *Presidents Given Immunity from Suits*, WASH. POST A6 (June 25, 1982), in 81 POWELL PAPERS, *supra* note 255, at 235 (same).

²⁵⁸ *Harlow v. Fitzgerald*, 457 U.S. 800, 816 & n.27 (1982). Indeed, Justice Powell explicitly mentioned this to his law clerk Richard Fallon during the drafting process of *Harlow*. See Memorandum from Lewis F. Powell, Jr.,

The objectifying cases as a whole, too, followed a largely *Marburian* trajectory. They gestured towards the need for stronger enforcement of constitutional rights against state officials by imposing a second prong in *Wood*, before drawing back in the opposite direction by making the objective prong more permissive in *Navarette* and dropping the subjective one in *Harlow*.

But while the trend towards objectification was fairly linear and followed a *Marburian* trajectory, a second trend went much less smoothly. This was Justice White's attempt to create a uniform standard by insisting on a functional analysis when it came to deciding whether an official received qualified or absolute immunity.

3. Towards a Uniform Standard: "My Vote or Byron's!"

Harlow's main doctrinal innovation was to make the qualified-immunity standard objective. But as Justice Brennan made sure to underscore in his concurrence, a second important outcome was to establish that standard "across the board, to all government officials performing discretionary functions" other than the President.²⁵⁹ Such a broad pronouncement was by no means required in a case that involved only high-ranking advisors to the President. *Scheuer v. Rhodes* had made qualified immunity available to executive officers "in varying scope, . . . the variation being [in part] dependent upon the scope of discretion and

Assoc. Justice, U.S. Supreme Court, to Richard H. Fallon, Law Clerk, U.S. Supreme Court (Mar. 1, 1982), in 84 POWELL PAPERS, *supra* note 206, at 3 ("[Justice White] agrees now that virtually every plaintiff can make a jury case by alleging malice – a subjective issue of an officer's good faith. . . . [But] this should be a question of law for the court to decide."). See also Chen, *The Facts*, *supra* note 91, at 232, 262-70.

²⁵⁹ *Harlow*, 457 U.S. at 821 (Brennan, J., concurring) (internal quotation marks omitted).

responsibilities of the office.”²⁶⁰ This “sliding scale”²⁶¹ was now replaced by a uniform standard. This uniformization marked a second victory for Justice White. For years, he had fought to enshrine a *functional* approach when it came to determining whether an executive official was entitled to absolute or qualified immunity. Here, too, Justice White pursued a *Marburian* path, albeit with some setbacks.

A. *Imbler v. Pachtman and Butz v. Economou: Settling the Functional Approach*

Justice White developed the functional approach in a quick succession of opinions between 1975 and 1978. He argued that the common law had long recognized policy concerns that justified absolute immunity for legislative and judicial acts.²⁶² But when it came to immunizing executive functions, the common law invoked a different policy analysis, and was more equivocal on whether immunity was available.²⁶³ In making the initial determination of whether an official could invoke absolute or qualified immunity, the Court thus looked to the nature of the official’s challenged acts.²⁶⁴ For acts that were legislative or judicial in nature—*e.g.*, initiating prosecution,²⁶⁵ presiding over an administrative adjudication,²⁶⁶ or presenting

²⁶⁰ 416 U.S. 232, 247 (1974).

²⁶¹ Memorandum from Byron R. White to the Conference of the U.S. Supreme Court on *Kissinger v. Halperin* 30 (recirculated Apr. 9, 1981), in 84 POWELL PAPERS, *supra* note 206, at 61 (“*Scheuer* does indeed create a sliding scale in the scope of official immunity.”).

²⁶² See, *e.g.*, *Butz v. Economou*, 438 U.S. 478, 508-12 (1978) (White, J., majority opinion); *Stump v. Sparkman*, 435 U.S. 349, 356-58 (1976) (White, J., majority opinion); *Imbler v. Pachtman*, 424 U.S. 409, 434-35 (1976) (White, J., concurring in the judgment).

²⁶³ *Butz*, 438 U.S. at 504-08; *Imbler*, 424 U.S. at 433 (White, J., concurring in the judgment); *Wood v. Strickland*, 420 U.S. 308, 320-22 (1975) (White, J., majority opinion).

²⁶⁴ *Butz*, 438 U.S. at 508-10; *Imbler*, 424 U.S. at 433-34 (White, J., concurring in the judgment).

²⁶⁵ *Imbler*, 424 U.S. at 441 (White, J., concurring in the judgment).

²⁶⁶ *Butz*, 438 U.S. 513-16.

evidence in court²⁶⁷—, the officer would receive absolute immunity. For acts that were executive—*e.g.*, gathering evidence or other investigative activities—, immunity would be qualified.²⁶⁸ Justice White’s functional analysis thus ignored the branch of government to which the official technically belonged, as well as the rank that she held in that branch.²⁶⁹

Justice White’s focus on the official’s function over her rank and branch membership paved the way for a uniform qualified-immunity standard. Determining whether an official’s actions were executive or non-executive in nature in order to assign them qualified or absolute immunity was intricate enough. Distinguishing *among* executive acts and tailoring qualified-immunity standards to them would have been a step too far for the Court. *Scheuer*’s sliding scale was thus recast into a binary in *Butz* and *Harlow*. Qualified immunity was assigned to officials who performed executive functions, while absolute immunity was assigned to officials who performed legislative, judicial, or quasi-judicial functions.²⁷⁰ Justice White had laid the foundation for this analysis in his majority opinions in *Stump v. Sparkman*²⁷¹ and *Butz v. Economou*,²⁷² as well as his concurrence in *Imbler v. Pachtman*.²⁷³ The analysis in all of these opinions ended with the decision of whether the immunity accorded should be absolute or qualified. When *Butz* made the functional approach a requirement for determining whether

²⁶⁷ *Id.* at 516-18.

²⁶⁸ *Id.* at 508-10.

²⁶⁹ *Id.* at 512.

²⁷⁰ *Butz*, 438 U.S. at 504-08.

²⁷¹ 435 U.S. 349 (1976).

²⁷² 438 U.S. 478 (1978).

²⁷³ 424 U.S. 409 (1976).

qualified or absolute immunity applied, it set the stage for *Harlow*'s standard that applied "across the board" for all federal officials²⁷⁴—and soon to state officials as well.²⁷⁵

However, *Butz* was a hard-fought case, decided by a bare majority. In fact, Court papers on *Butz* published by the Burger Court Opinion Writing Database suggest that Justice Rehnquist—who wanted absolute immunity for all of the officials involved in that case²⁷⁶—initially commanded a majority.²⁷⁷ As Justice White convinced a wavering Justice Powell to join his side,²⁷⁸ Justice Rehnquist announced that he would “now . . . convert[] [his] initial memorandum . . . into a dissent.”²⁷⁹ In caustic language, Rehnquist made his position clear:

My biggest concern . . . is not with the illogic or impracticality of today's decision, but rather with the potential for disruption of Government that it invites. The steady increase in litigation, much of it directed against governmental officials and virtually all of which could be framed in constitutional terms, cannot escape the notice of even the most casual observer.²⁸⁰

For Rehnquist, attempting to “hold a middle ground” was between vindicating constitutional rights and ensuring vigorous exercise of governmental functions was admirable, but ultimately

²⁷⁴ *Harlow v. Fitzgerald*, 457 U.S. 800, 818 n.30 (1982) (citing *Butz*, 438 U.S. at 504).

²⁷⁵ See Crocker, *supra* note 94, at 1432-33 (discussing the ambiguity of when exactly the Court seems to have decided that the same qualified-immunity standard applies to federal and state officials).

²⁷⁶ *Butz*, 438 U.S. at 517-18, 525-30 (Rehnquist, J., dissenting).

²⁷⁷ See Note from Byron R. White, Assoc. Justice, U.S. Supreme Court, to William H. Rehnquist, Assoc. Justice, U.S. Supreme Court, (Jan. 5, 1978), in Paul J. Wahlbeck et al., *Butz v. Economou*, 438 U.S. 478 (1978), THE BURGER COURT OPINION WRITING DATABASE 12 (telling Justice Rehnquist he would “circulate a memorandum . . . dissenting in part with the memorandum you have circulated”), http://supremecourttopinions.wustl.edu/files/opinion_pdfs/1977/76-709.pdf [hereinafter *Butz Papers*].

²⁷⁸ Compare Note from Lewis F. Powell, Jr., Assoc. Justice, U.S. Supreme Court, to William H. Rehnquist, Assoc. Justice, U.S. Supreme Court, (Jan. 6, 1978) (expressing agreement with Rehnquist's judgment, but tentative agreement with White's reasoning) with Note from Lewis F. Powell, Jr., Assoc. Justice, U.S. Supreme Court, to Byron R. White, Assoc. Justice, U.S. Supreme Court, (May 27, 1978) (confirming that he wants to join White's opinion), *Butz Papers*, *supra* note 277, at 27, 29.

²⁷⁹ Note from William H. Rehnquist, Assoc. Justice, U.S. Supreme Court, to Byron R. White, Assoc. Justice, U.S. Supreme Court, (May 22, 1978), *Butz Papers*, *supra* note 277, at 42.

²⁸⁰ 438 U.S. at 526 (Rehnquist, J., dissenting).

impossible.²⁸¹ The former simply had to give, and he left little doubt that the last word on the matter had not yet been spoken.²⁸²

White's victory for the functional approach meant a victory for qualified immunity, a middle position between the conservative Justices' preference for absolute immunity and § 1983's strict liability scheme. With *Butz* on the books, the stage was set for the landmark twin cases *Harlow* and *Nixon v. Fitzgerald*. Qualified immunity would become the standard for executive officials across the board—except for the President of the United States. But the battles behind the scenes suggested that the Court would soon move away from the *Marburian* wisdom that qualified immunity embodied. *Harlow v. Fitzgerald* cannot be understood without the fractious disagreements that engulfed the Court in *Nixon v. Fitzgerald*. It is thus worth analyzing *Nixon* in some detail.

B. *Nixon v. Fitzgerald: Qualified or Absolute Immunity for a Controversial President?*

The disagreement over presidential immunity reached back into the previous term, when an eight-member Court had sparred over the question of immunity for the President and his closest aides in *Kissinger v. Halperin*.²⁸³ Justice Powell had been tasked to write for a five-member majority that favored immunizing the President—either on the constitutional theory that separation of powers required it or on the grounds that no *Bivens* action could lie against

²⁸¹ *Id.* at 530.

²⁸² *Id.* In *Harlow*, too, Rehnquist made his convictions clear. While he felt compelled to join the majority with *Butz* on the books, he advertised that he would a “reexamine [the] holding of *Butz* . . . with alacrity” if that opportunity presented itself. 457 U.S. at 822 (Rehnquist, J., concurring). See *infra*, Section III.B.i., for Chief Justice Rehnquist's efforts to recalibrate qualified immunity as more forgiving for government officials.

²⁸³ 452 U.S. 713 (1981). Justice Rehnquist had recused himself because one of the parties was his former boss, Attorney General John Mitchell. Memorandum from William H. Rehnquist, Assoc. Justice, U.S. Supreme Court, to the Conference of the U.S. Supreme Court (May 27, 1981), in *Halperin Papers*, *supra* note 244, at 98.

a President.²⁸⁴ Powell's opponent in the debate was Justice White, who tirelessly argued that both the President and his aides should only receive qualified immunity based on his functional approach and called absolute immunity for the President and his aides "gross overkill."²⁸⁵ When Justice Marshall defected from the majority at the last moment,²⁸⁶ the Court was deadlocked four-to-four.²⁸⁷ Powell authored a short *per curiam* opinion that left the lower court's denial of absolute immunity for the President in place.²⁸⁸ The Court opted to hear the *Fitzgerald* cases for next term, in which all nine Justices could participate.

The facts in *Harlow* and *Nixon* were, again, highly political: Arthur Fitzgerald, a contractor with the U.S. Air Force under President Nixon, had blown the whistle on billions of dollars of wasteful government spending in 1969.²⁸⁹ After testifying before Congress, Fitzgerald was moved to another government job and ultimately fired.²⁹⁰ He brought a *Bivens* suit for wrongful termination suit against the President and several of his aides.²⁹¹ The case came to the Court in 1982, after it had already decided the highly contentious Nixon tapes

²⁸⁴ Memorandum from Lewis F. Powell, Jr., Assoc. Justice, U.S. Supreme Court, to the Conference of the U.S. Supreme Court (Mar. 18, 1981), in *Halperin Papers*, *supra* note 244, at 72-73.

²⁸⁵ Letter from Byron R. White, Assoc. Justice, U.S. Supreme Court, to Lewis F. Powell, Jr., Assoc. Justice, U.S. Supreme Court (Mar. 19, 1981), in *Halperin Papers*, *supra* note 244, at 23-24.

²⁸⁶ Memorandum from Thurgood Marshall, Assoc. Justice, U.S. Supreme Court, to the Conference of the U.S. Supreme Court (May 27, 1981), in *Halperin Papers*, *supra* note 244, at 33.

²⁸⁷ Letter from Lewis F. Powell, Jr., Assoc. Justice, U.S. Supreme Court, to Thurgood Marshall, Assoc. Justice, U.S. Supreme Court (Mar. 19, 1981), in *Halperin Papers*, *supra* note 244, at 75-76.

²⁸⁸ 452 U.S. 713 (1981).

²⁸⁹ *Nixon v. Fitzgerald*, 457 U.S. 731, 733-34 (1982).

²⁹⁰ *Id.* at 735-36.

²⁹¹ *Id.* at 739-41. While *Nixon* and *Harlow* were *Bivens* suits, the immunity standard announced in *Harlow* was intended for suits against state officials under § 1983 as well, and would quickly be adopted for those in *Davis v. Scherer*, 468 U.S. 183 (1984). See also Crocker, *supra* note 94, at 1432-33.

case,²⁹² the Nixon wiretaps case,²⁹³ and the *Halperin* case.²⁹⁴ The two *Fitzgerald* cases capped off a spate of cases in which the executive was brought before the nation's highest tribunal.²⁹⁵ From an institutional perspective, it thus seemed wise for the Court to tread lightly, and to speak with a single voice.

But divisions were no less sharp this time around.²⁹⁶ Chief Justice Burger in particular did not mince his words when he assigned the opinions to Justice Powell: “[T]he assignment is made to you but I feel obligated to state that my view is irrevocable on absolute immunity. . . . I am still unable to understand why we should ‘duck’ that issue when the votes are there. . . . Put ‘hard,’ your choice is my vote or Byron’s!”²⁹⁷ Justice Powell, trying to convince the Chief Justice that “in this case particularly – [the Court had] a strong institutional reason for avoiding fractionalization,”²⁹⁸ spent the rest of the term attempting to get “a [full] Court opinion, as well as as many votes for the judgment as can be mustered.”²⁹⁹ Chief Justice Burger would get his wish on absolute immunity for the President in *Nixon v. Fitzgerald*. But the key figure would

²⁹² *United States v. Nixon*, 418 U.S. 683 (1974).

²⁹³ *United States v. United States District Court (Keith)*, 407 U.S. 297 (1972).

²⁹⁴ *Halperin v. Kissinger*, 606 F.2d 1192 (D.C. Cir. 1979), *aff'd by an equally divided Court*, 452 U.S. 713 (1981) (per curiam). Morton Halperin had brought a damages suit against President Nixon, National Security Adviser Henry Kissinger, Attorney General John Mitchell, and White House Chief of Staff H.R. Haldeman.

²⁹⁵ For an excellent summary of how the Burger Court navigated these confrontations with the presidency, see GRAETZ & GREENHOUSE, *supra* note 163, at 297-338.

²⁹⁶ In fact, the Court was even more fractured now because of a contingent settlement agreement that appeared after the Court had granted certiorari. At conference, Justices Brennan and Blackmun thus wanted to deny certiorari as improvidently granted. Lewis F. Powell, Jr., Notes on Second Conference, *Nixon v. Fitzgerald* (No. 79-1738), (Dec. 14, 1981), in 81 POWELL PAPERS, *supra* note 255, at 90-91. For more details on the Court's fractured votes, see *supra* note 207 and accompanying text.

²⁹⁷ Draft Letter from Warren E. Burger, Chief Justice, U.S. Supreme Court, to Lewis F. Powell, Jr., Assoc. Justice, U.S. Supreme Court (Dec. 14, 1981), in 81 POWELL PAPERS, *supra* note 255, at 93.

²⁹⁸ Draft Letter from Lewis F. Powell, Jr., Assoc. Justice, U.S. Supreme Court, to Warren E. Burger, Chief Justice, U.S. Supreme Court (Dec. 16, 1981), in 81 POWELL PAPERS, *supra* note 255, at 120.

²⁹⁹ *Id.* at 116.

once again be Justice White, who conclusively enshrined the objective standard and the functional approach in *Harlow*. While *Harlow* managed to adhere to *Marbury*'s wisdom, *Nixon* would cast it to the winds.

C. *Ignoring Marbury's Wisdom: Absolute Immunity for the President*

In conference for the *Nixon* case, three different positions emerged on the merits.³⁰⁰ Justices Marshall and White favored a narrow holding that *Bivens* suits for retaliatory termination did not lie against the President.³⁰¹ Justices Powell and Stevens preferred to go broader and hold that no *Bivens* suit could lie against the President, but leaving open the question of whether Congress could impose liability on the President by statute.³⁰² The Chief Justice and Justices Rehnquist and O'Connor sought to bar *any* civil suit against the President for official acts done in office (statutory or implied), and voted for absolute immunity.³⁰³ While all three holdings would have bristled against the *Marburian* principle of recognizing rights but limiting their remedies, there was a wide gulf between Justice White's position foreclosing a single cause of action against the President and the Chief Justice's barring of all suits.

Justice Powell tried to corral a majority around the broad no-*Bivens* rationale. To him, the result would be virtually identical to finding absolute immunity, as he could not imagine

³⁰⁰ Justices Brennan and Blackmun voted to dismiss certiorari as improvidently granted when they found out that a contingent settlement agreement between Nixon and Fitzgerald may have mooted the case. See 81 POWELL PAPERS *supra* note 255.

³⁰¹ Lewis F. Powell, Jr., Notes on Second Conference, *Nixon v. Fitzgerald* (No. 79-1738), (Dec. 14, 1981), in 81 POWELL PAPERS, *supra* note 255, at 90-92.

³⁰² *Id.*

³⁰³ *Id.*

that Congress would attempt to impose statutory liability on the President.³⁰⁴ By the same token, Justices White and Marshall could concur in the judgment on the narrow no-*Bivens* rationale,³⁰⁵ while the only dissent would be that certiorari had been improvidently granted.³⁰⁶ The Court could have essentially spoken with a single voice in a case involving “a highly controversial president,”³⁰⁷ without straying too far from the *Marburian* balance. But Chief Justice Burger and Justice Rehnquist drove a hard line, insisting that absolute immunity was the threshold issue.³⁰⁸ In order to avoid a fractured Court without a majority opinion, Justices Powell and Stevens abandoned their idea to reason with the Chief,³⁰⁹ and instead decided to join the absolute-immunity bloc.³¹⁰ This was far from ideal for Justice Powell, who worried about the partisan optics of a bare majority giving absolute immunity to the President where “each of the five [majority Justices had] been appointed by Republicans.”³¹¹ Burger, Powell, and Rehnquist had been appointed by Nixon, the very president who was a party in this suit;

³⁰⁴ Draft Letter from Lewis F. Powell, Jr., Assoc. Justice, U.S. Supreme Court, to Warren E. Burger, Chief Justice, U.S. Supreme Court (Dec. 16, 1981), in 81 POWELL PAPERS, *supra* note 255, at 119.

³⁰⁵ *Id.*

³⁰⁶ *Id.*

³⁰⁷ *Id.* at 116.

³⁰⁸ See, e.g., Note from Lewis F. Powell, Jr., Assoc. Justice, U.S. Supreme Court, to self (Dec. 14-15, 1981) (“I used these notes – to no avail – in trying to persuade the CJ + WHR to go with ‘no cause of action.’”), in 81 POWELL PAPERS, *supra* note 255, at 94.

³⁰⁹ Draft Letter from Lewis F. Powell, Jr., Assoc. Justice, U.S. Supreme Court, to Warren E. Burger, Chief Justice, U.S. Supreme Court (Dec. 16, 1981) (“not sent – see my letter to CJ of [Dec. 17, 1981]”), in 81 POWELL PAPERS, *supra* note 255, at 96.

³¹⁰ Letter from Lewis F. Powell, Jr., Assoc. Justice, U.S. Supreme Court, to Warren E. Burger, Chief Justice, U.S. Supreme Court (Dec. 17, 1981), in 81 POWELL PAPERS, *supra* note 255, at 126-27.

³¹¹ Personal Notes of Lewis F. Powell, Jr., Assoc. Justice, U.S. Supreme Court (Dec. 14-15, 1981) (observing that, if he and Justice Stevens were to join the Chief Justice and Justices Rehnquist and O’Connor in their absolute-immunity rationale, “the vote for Nixon would be 5 – 4 — with each of the five having been appointed by Republican[s]”), in 81 POWELL PAPERS, *supra* note 255, at 94.

Stevens had been appointed by Ford, who had pardoned Nixon;³¹² and O'Connor was a Reagan-appointee.

Moreover, this hard line predictably irked Justice White, who had moved away from his initial preference for *qualified* immunity for the President to settle on the no-*Bivens* rationale in conference.³¹³ White now set out to write a vociferous four-Justice dissent that provided quotable lines to the press that the majority put the President “above the law.”³¹⁴ He also pried open an ambiguity that Justices Powell and Stevens would have preferred to leave unaddressed: if the majority held the President to be absolutely immune from suit on constitutional grounds, the implication was not only that the President was exempted from *Bivens* suits, but also that Congress could not create statutory causes of action against the President if it chose to do so.³¹⁵ Justice Stevens had only agreed to vote for the absolute immunity position on the condition that the issue of Congressional liability would not be explicitly addressed.³¹⁶ When Justice White’s draft dissents pointed out the dodge, Justice Powell had to mediate between Justice Stevens and Chief Justice Burger, who withheld his

³¹² *Ford Pardons Nixon*, HISTORY (Nov. 24, 2009), <https://www.history.com/this-day-in-history/ford-pardons-nixon>.

³¹³ Lewis F. Powell, Jr., Notes on Second Conference, *Nixon v. Fitzgerald* (No. 79-1738), (Dec. 14, 1981), in 81 POWELL PAPERS, *supra* note 255, at 90.

³¹⁴ Compare *Nixon v. Fitzgerald*, 457 U.S. 731, 766 (1982) (White, J., dissenting), with Linda Greenhouse, *High Court Holds President Immune from Damage Suits*, N.Y. TIMES A12 (June 25, 1982), in 81 POWELL PAPERS, *supra* note 255, at 239, and Fred Barbash, *Presidents Given Immunity from Suits*, WASH. POST A6 (June 25, 1982), in 81 POWELL PAPERS, *supra* note 255, at 235.

³¹⁵ *Nixon*, 457 U.S. at 765 (White, J., dissenting). Justice Blackmun made the points even more emphatically in his dissent, *id.* at 797-98 (Blackmun, J., dissenting).

³¹⁶ Letter from Lewis F. Powell, Jr., Assoc. Justice, U.S. Supreme Court, to Warren E. Burger, Chief Justice, U.S. Supreme Court (Dec. 17, 1981), in 81 POWELL PAPERS, *supra* note 255, at 127.

vote and threatened to reduce the majority opinion to a plurality.³¹⁷ Powell, beseeching the Chief that “[a] plurality on an issue as inflammatory as this one . . . [would] invite future challenges when the composition of the Court changes,”³¹⁸ was finally able to win the Chief over without having to foreclose explicit statutory liability.³¹⁹

This standoff between White and Burger was rooted in a deep philosophical disagreement over which basic analytical approach was appropriate in the immunity context. Burger had wanted absolute immunity for the President on constitutional grounds *and* for his aides on a theory of derivative immunity.³²⁰ This threatened to create a doctrinal conflict with the functional approach that White had established in *Butz* just four years earlier. In essence, the Chief Justice wanted to abandon the *Marburian* middle ground when it came to the President and high-ranking officials by foreclosing any cause of action against them. Justice White was committed to holding the middle ground of recognizing rights, but reserving discretion on the remedy. By contrast, *Harlow* marked a victory for Justice White’s middle position. *Harlow* and *Nixon* thus present a clear contrast of when the Court managed to hew close to *Marbury*’s wisdom, and when it did not.

4. Summing Up: Qualified Immunity’s *Marburian* Origins

³¹⁷ See Letter from Warren E. Burger, Chief Justice, U.S. Supreme Court, to Lewis F. Powell, Jr., Assoc. Justice, U.S. Supreme Court (June 8, 1982), *in* 81 POWELL PAPERS, *supra* note 255, at 171 (discussing his concurrence in the judgment and urging Powell to “bite the bullet” and not reserve the question of statutory liability).

³¹⁸ Letter from Lewis F. Powell, Jr., Assoc. Justice, U.S. Supreme Court, to Warren E. Burger, Chief Justice, U.S. Supreme Court (June 8, 1982), *in* 81 POWELL PAPERS, *supra* note 255, at 172-73.

³¹⁹ The Chief joined on June 11, 1982. Personal Note Chart of Lewis F. Powell, Jr., Assoc. Justice, U.S. Supreme Court, *in* 81 POWELL PAPERS, *supra* note 255, at 229. For the final formulation, see *Nixon*, 457 U.S. at 748 n.27 (“[O]ur holding today need only be that the President is absolutely immune from civil damages liability for his official acts in the absence of explicit affirmative action by Congress.”).

³²⁰ *Harlow v. Fitzgerald*, 457 U.S. 800, 822-29 (1982) (Burger, C.J., dissenting).

The long first decade of qualified immunity doctrine, from *Pierson v. Ray* to *Harlow v. Fitzgerald*, was a modern *Marburian* move. While the “Congress [of 1871 had] recognized the need for original federal jurisdiction as a means to provide [some] federal control over the unconstitutional acts of state officials,”³²¹ the political realities of the 1970s required a partial “high-minded refus[al].”³²² The Court threaded the needle by standing for the proposition that, while the “very essence of civil liberty certainly consists in the right of every individual to claim the protection of the laws,”³²³ rights and remedies do not necessarily coincide when it comes to suits against a government official.³²⁴ Individuals should be able to bring claims, but *courts* should have a significant word in determining whether such claims should go to trial. Seemingly refusing the Congressional grant of jurisdiction over constitutional and statutory claims, the Court created and claimed a different judicial power: deciding to decide when to hear those claims.

Indeed, by charting such a middle course by refusing one power but claiming another, and by preserving a cause of action but rationing the remedy,³²⁵ the Court avoided aligning itself with either side in cases that were often highly political. *Pierson v. Ray* pitted police whose

³²¹ *District of Columbia v. Carter*, 409 U.S. 418, 428 (1973).

³²² MCCLOSKEY, *supra* note 113, at 26.

³²³ *Butz v. Economou*, 438 U.S. 478, 485 (1978) (quoting *Marbury v. Madison*, 5 U.S. (1 Cranch) 137, 163 (1803)). See also *Nixon v. Fitzgerald*, 457 U.S. 731, 783, 789, 797 (1982) (White, J., dissenting) (quoting the same language three separate times). Indeed, White charged that giving the President absolute immunity abandoned *Marbury*’s fundamental principle. *Id.* at 768 (White, J., dissenting) (criticizing that “if the laws furnish no remedy for the violation of a vested legal right” the government can no longer be “termed a government of laws, and not of men” (quoting *Marbury*, 5 U.S. at 163)). The majority laconically responded that “*Marbury* does not establish that the individual’s protection must come in the form of a particular remedy. *Marbury*, it should be remembered, *lost* his case in the Supreme Court.” *Id.* at 754 n.37 (majority opinion).

³²⁴ See Jeffries, *Right-Remedy Gap*, *supra* note 105; Fallon, *Bidding Farewell*, *supra* note 70; Amar, *supra* note 112; Huq, *supra* note 112.

³²⁵ Huq, *supra* note 112.

“real aim was to preserve segregation”³²⁶ against activists “provok[ing] a breach of the peace in the form of violence.”³²⁷ *Scheuer v. Rhodes* featured competing narratives of an “insurrection and mob rule”³²⁸ against “inadequately trained and incapable troops [who engaged in] unlawful shootings.”³²⁹ The subtext of *Wood v. Strickland* was a shift in disciplinary thinking in the school context, especially for rebellious girls flaunting traditional standards of morality. And *Harlow v. Fitzgerald* featured the aides of a “highly controversial President”³³⁰ against a whistleblower who undermined government business and failed to understand that “loyalty was the name of the game.”³³¹ (*Nixon*, as noted above, was the exception that proved the rule underlying *Marbury*’s wisdom.³³²)

Indeed, all of those cases arose in contexts of larger struggles between “the people” and the government: the Freedom Rides, student protests against the Vietnam War, initiatives to introduce due process protections in public schools, efforts to combat a rising prison population in the incipient War on Crime,³³³ and (of course) President Nixon’s “rogue government [that] had mounted a campaign of extralegal and illegal activities.”³³⁴ At a time

³²⁶ *Pierson Petitioner’s Brief*, *supra* note 122, at 17.

³²⁷ *Pierson Respondent’s Brief*, *supra* note 148, at 18.

³²⁸ *Krause v. Rhodes*, 471 F.2d 430, 435 (6th Cir. 1972); *Krause Scheuer Petitioner’s Brief*, *supra* note 178, at 19.

³²⁹ Brief for Petitioner at 6-7, *Scheuer v. Rhodes*, 416 U.S. 232 (1974) (No. 72-914), 1973 WL 171771, at *8.

³³⁰ Draft Letter from Lewis F. Powell, Jr., Assoc. Justice, U.S. Supreme Court, to Warren E. Burger, Chief Justice, U.S. Supreme Court (Dec. 16, 1981), *in* 81 POWELL PAPERS, *supra* note 255, at 116.

³³¹ *Nixon v. Fitzgerald*, 457 U.S. 731, 736 (1982) (quoting a memorandum of presidential aide Alexander Butterfield recommending not to reemploy Fitzgerald for a while).

³³² See *supra* note [X] and accompanying text.

³³³ See *Procunier v. Nararette*, 434 U.S. 555, 557-60 (1978).

³³⁴ SCHULMAN, *supra* note 175, at 43-44.

when the caseload of federal civil rights cases had gone from 296 in 1961 to 13,113 in 1977,³³⁵ and when eroding trust in government and public institutions led to a crisis of authority in the United States,³³⁶ the Supreme Court made its best efforts not to take sides.

III. LOSING SIGHT OF *MARBURY*: THE DOOMED PROJECT OF *SAUCIER V. KATZ*

The Court had carved out a new standard in *Harlow v. Fitzgerald* and turned the page on many of the questions it had debated for years. Qualified immunity was to be a standard that applied equally to all officers performing executive functions (federal and state). It was objective. It struck a balance between vindicating constitutional rights and giving officers reasonable room for error in the execution of their duties. And ideally (but not necessarily), it would resolve the immunity question before discovery.³³⁷

Yet, those questions continued to reappear in two new forms in the years following *Harlow*. First, the Court had to clarify the meaning of “clearly established law of which a reasonable official would have known,” especially when the underlying constitutional right was also determined under a standard of reasonableness.³³⁸ Second, the Court faced the conundrum of how to sequence and distinguish the immunity analysis from the analysis of the underlying constitutional claim.³³⁹ In both arenas, the Court would slowly abandon the

³³⁵ Butz v. Economou, 438 U.S. 478, 526 (1978) (Rehnquist, J., dissenting) (citing statistic of the Director of the Administrative Office of the United States Courts). *But see* Eisenberg, *supra* note 230, at 484 (presenting, as early as 1982, empirical indicating that “section 1983 cases are not overwhelming the federal courts”).

³³⁶ SCHULMAN, *supra* note 175, at 9-10, 48, 51, 147.

³³⁷ See *supra* Sections II.B–C.

³³⁸ See *infra* Section III.B.ii.

³³⁹ See *infra* Section III.B.i.

institutional considerations underlying the *Marburian* move by restricting access to individual rights and chipping away at its own discretion on the remedies side.³⁴⁰

But the Court’s qualified immunity jurisprudence between *Harlow* and *Saucier* was not a straightforward departure from *Marbury*. It was complex and full of competing cross-currents. Before turning to the core issues of the meaning of “clearly established” and of sequencing the merits and immunity analyses, it is appropriate to briefly analyze the side battles that were fought around qualified immunity. This Part will demonstrate that *those* cases closely adhered to the Court’s reasoning in *Marbury*, underscoring just how out of character it was for the Court to depart from its *Marburian* wisdom when it came to deciding the substance and analytical structure of the qualified-immunity standard.

1. *Marburian* Moves in Peripheral Questions

Between 1982 and 2001, the Supreme Court worked out several questions in qualified immunity that did not deal with the substantive standard or its analytical structure. This involved refining the functional determination of whether absolute or qualified immunity applied, the appealability of qualified-immunity denials, the doctrine’s relationship to pleading requirements in unconstitutional motive cases, the doctrine’s relationship to defenses in 18 U.S.C. § 242 (§ 1983’s criminal counterpart), and the availability of qualified immunity to private actors sued under § 1983. In all of them, the Court continued to adhere to *Marbury*’s wisdom.

A. *Malley v. Briggs and Progeny: Continuing the Functional Approach*

³⁴⁰ See *infra* Sections III.B.i–iii.

When it came to determining whether an official should receive absolute or qualified immunity, the functional approach was alive and well in the post-*Harlow* era. Justice White continued to be its main driver.³⁴¹ The underlying analytical approach was the same in every opinion: Section 1983 did not intend to abrogate common-law immunities. Since the common law was less conclusive on executive immunity than it was on legislative or judicial immunity, the norm for executive officials was qualified immunity. Only a strong common-law tradition to the contrary could warrant a grant of absolute immunity.³⁴² In *Mitchell v. Forsyth*, this common-law-based analysis foreclosed the rationale that the Attorney General would receive absolute immunity “for acts performed in the exercise of his national security functions.”³⁴³ For the executive function of performing investigative acts, only qualified immunity was available.³⁴⁴ Judicial acts, by contrast, were entitled to absolute immunity.

The Court thus began to draw finer and finer lines to distinguish executive officers’ judicial functions from their executive functions. Police officers, for example, received absolute immunity for statements made as a witness in a judicial proceeding,³⁴⁵ but qualified immunity in their investigative function of applying for a warrant.³⁴⁶ Similarly, prosecutors received absolute immunity for their quasi-judicial function of appearing to testify and present

³⁴¹ See *infra*, notes 274–279.

³⁴² See *infra*, notes 274–279.

³⁴³ 472 U.S. 511, 519-20 (majority opinion by White, J.).

³⁴⁴ See *id.* at 519-24 for standard functional analysis.

³⁴⁵ *Briscoe v. LaHue*, 460 U.S. 325, 345-46 (1983) (majority opinion by White, J.). See *id.* at 330-36 and 341-46 for standard functional analysis.

³⁴⁶ *Malley v. Briggs*, 475 U.S. 335, 344 (1986) (majority opinion by White, J.). See *id.* at 339-45 for standard functional analysis.

evidence in a probable-cause hearing,³⁴⁷ but not for giving legal advice to police trying to establish probable cause,³⁴⁸ for acts of gathering evidence before probable cause was established,³⁴⁹ or making statements to the press.³⁵⁰

Perhaps most interestingly, the Court denied *any* immunity to privately employed prison guards on the theory that its functional analysis only applied to government officials.³⁵¹ Performing the same tasks, but beyond the aegis of government control and supervision, private prisons did not give rise to the same common-law and policy balancing concerns as those directly administered by the government.³⁵² When the government was not involved, no *Marburian* move was necessary.

The functional post-*Harlow* cases thus continued to pursue the same middle ground as before. Drawing the line between judicial and executive acts was not a matter of “a freewheeling policy choice,” but an exercise in the faithful interpretation of the congressional

³⁴⁷ *Burns v. Reed*, 500 U.S. 478, 492 (1991) (majority opinion by White, J.). See *id.* at 484-96 for standard functional analysis.

³⁴⁸ *Id.* at 496.

³⁴⁹ *Buckley v. Fitzsimmons*, 509 U.S. 259, 275 (1993) (majority opinion by Stevens, J.). See *id.* at 268-71 for standard functional analysis.

³⁵⁰ *Id.* at 277.

³⁵¹ *Richardson v. McKnight*, 521 U.S. 399, 407-08, 412 (1997) (majority opinion by Breyer, J.). Fifteen years later, however, the Court decided to cabin *Richardson* in *Filarsky v. Delia*, 566 U.S. 377 (2012) (holding that qualified immunity is available to a private individual temporarily retained by the government to carry out its work).

³⁵² *Id.* at 407-12.

intent of § 1983.³⁵³ It was a matter of ensuring—at least nominally³⁵⁴—that the qualified immunity standard would not slide into absolute immunity.

B. *Mitchell v. Forsyth and Progeny: Insisting on Questions of Law*

The Court also continued to protect the new judicial power it had created in *Harlow* by ensuring that § 1983 actions could be brought against a wide range of defendants,³⁵⁵ that the protection flowing from qualified immunity did not affect the pleading standards for the underlying constitutional merits case,³⁵⁶ and that, in suits for criminal violations of constitutional rights brought under 18 U.S.C. § 242, the statute’s notice standard was no stricter than what § 1983 required.³⁵⁷ The decisions, in other words, ensured that litigants could reach the doorstep of the courthouse, but left it up to the judiciary to decide who could enter.

Moreover, the Court gave appellate courts wide latitude over legal determinations in the qualified immunity analysis. *Mitchell v. Forsyth* expanded the collateral order doctrine to make denials of qualified immunity on summary judgment immediately appealable.³⁵⁸ *Johnson*

³⁵³ *Malley*, 475 U.S. at 342; *Burns*, 500 U.S. at 493; *Buckley*, 509 U.S. at 268; *Tower v. Glover*, 467 U.S. 914, 922-23 (1984) (all quoting the same language). This language contrasted starkly with Justice Scalia’s open departure from “slavishly deriv[ing] immunity rules] from the often arcane rules of the common law,” *Anderson v. Creighton*, 483 U.S. 635, 645 (1987), or Justice Kennedy’s creation of the *Saucier* two-step. See *infra* Section III.B.iii.

³⁵⁴ The real crux was, of course, that the qualified-immunity standard was beginning to move closer and closer to absolute immunity. *Infra*, Part IV.

³⁵⁵ *Hafer v. Melo*, 502 U.S. 21, 27-28 (1991) (clarifying that state officials sued in their official *capacities* can be held *personally* liable in damages); *Wyatt v. Cole*, 504 U.S. 158, 162 (1992) (holding that § 1983 suits can sometimes be brought against non-governmental actors); *Richardson v. McKnight*, 521 U.S. 399, 402-07, 412 (1997) (denying immunity to privately employed prison officials while bringing them under the purview of § 1983).

³⁵⁶ *Crawford-El v. Britton*, 523 U.S. 574, 584-93 (1998) (insisting that the *Harlow* standard did not raise the burden of proof in unconstitutional motive cases). “The reasoning in *Harlow*, like its specific holding, does not justify a rule that places the thumb on the defendant’s side of the scales when the merits of the claim that a defendant knowingly violated the law are being resolved. . . . [T]o change the burden of proof for an entire category of claims would stray far from the traditional limits of judicial authority.” *Id.* at 593-94.

³⁵⁷ *United States v. Lanier*, 520 U.S. 259, 270-71 (1997) (holding that the “fair warning” requirement of 18 U.S.C. § 242 was no higher than the “clearly established” standard of 42 U.S.C. § 1983).

³⁵⁸ 274 U.S. 511, 524-30 (1985).

v. Jones clarified that those denials could only be immediately appealed if they were based on a legal determination (such as that the right had not been clearly established), *not* on a determinations that the complaints raised genuine issues of fact.³⁵⁹ And *Elder v. Holloway* held that, in reviewing whether a right was clearly established, an appellate court could consider the universe of applicable caselaw de novo, even when plaintiff or the court below had failed to identify it.³⁶⁰ Here, too, the Court made sure that it and the appeals courts had maximal latitude in administering the standard, not juries or factfinders.³⁶¹

2. Two Departures from *Marbury*

In the twenty years after *Harlow*, many of the Supreme Court's qualified-immunity decisions continued to follow the *Marburian* policy balance of making § 1983 suits available to individual plaintiffs, carefully determining where qualified instead of absolute immunity was appropriate, and drawing discretion to itself in determining the appropriate remedy. The truly important issues, however, lay in spelling out the content and analytical structure of the new qualified-immunity standard. Here, a decidedly un-*Marburian* picture began to emerge. Again, two separate trends took shape. This time, it was the interpretation that "clearly established" required a higher level of specificity for the right in question than the constitutional merits standard, and the threshold question of whether the allegedly violated right existed at all. And just as in *Harlow*, these two developments would converge in a landmark decision: *Saucier v. Katz*.³⁶²

³⁵⁹ 515 U.S. 304, 313-18 (1995).

³⁶⁰ 510 U.S. 510, 512-14 (1994).

³⁶¹ See Chen, *The Facts*, *supra* note 91.

³⁶² 533 U.S. 194 (2001).

A. The Chief's Long Game: Questioning the Underlying Right

One of *Saucier*'s key holdings was that courts deciding a § 1983 or *Bivens* suit had to first determine whether plaintiff had sufficiently alleged a rights violation on the merits, and then, if necessary, proceed to the immunity analysis of whether that right was clearly established at the time of the challenged conduct.³⁶³

This sequencing requirement did not come out of thin air, but had been a long time in the making. Between *Harlow* and *Saucier*, the Court treated the question of whether a constitutional violation had been properly alleged as a threshold matter (and suggested to lower courts to do the same) in six cases: *Graham v. Connor*,³⁶⁴ *Siegert v. Gilley*,³⁶⁵ *Hunter v. Bryant*,³⁶⁶ *County of Sacramento v. Lewis*,³⁶⁷ *Conn v. Gabbert*,³⁶⁸ and *Wilson v. Layne*.³⁶⁹ Four of those opinions—*Graham*, *Siegert*, *Conn*, and *Wilson*—were authored by Chief Justice Rehnquist, and he explicitly concurred in *Lewis* to reiterate the proper sequencing of the analysis.³⁷⁰ Indeed, the proposition that “[t]he first inquiry in any § 1983 suit . . . is whether the plaintiff has been deprived of a [constitutional] right” goes all the way back to 1979, when then-Justice Rehnquist’s penned the majority opinion to *Baker v. McCollan*.³⁷¹ Justice Kennedy’s fixing of

³⁶³ 533 U.S. at 201.

³⁶⁴ 490 U.S. 386, 393 (1989).

³⁶⁵ 500 U.S. 226, 231 (1991).

³⁶⁶ 502 U.S. 228-29 (1991).

³⁶⁷ 523 U.S. 841 n.5 (1998).

³⁶⁸ 526 U.S. 286, 290 (1999).

³⁶⁹ 526 U.S. 603, 609 (1999).

³⁷⁰ 523 U.S. at 854-55.

³⁷¹ 443 U.S. 137, 140 (1979).

the proper sequence in *Saucier* had thus been a long time in the making. The architect had been none other than Chief Justice William Rehnquist.

At first blush, the *Marburian* implications for Rehnquist's threshold inquiry may not be clear. Facially, the inquiry could be *Marburian* if it simply required that a court identify and affirm the specific right that the individual sought to vindicate. But the anti-*Marburian* tendencies of the threshold analysis become clear from three observations. First, the inquiry was not *which* right plaintiff alleged to have been violated, but *whether* a rights violation had been alleged. Plaintiff had to make a showing that her claim survived this initial inquiry.³⁷²

Second, then-Justice Rehnquist had made no secret of his concern that § 1983 become “a font of tort law to be superimposed upon . . . the States” as early as 1976.³⁷³ Indeed, his *McCullan* opinion came down just one year after he had vociferously dissented from *Butz v. Economou*'s holding that all officials performing executive functions were to receive qualified immunity.³⁷⁴ Rehnquist ominously announced that “the all but inevitable result of [qualified immunity doctrine's attempt] to gain and hold a middle ground” would either result in “a significant impairment of . . . officials to carry out the[ir] duties” or in a “necessarily unprincipled and erratic judicial ‘screening’ of claims.”³⁷⁵

Third, out of the six post-*Harlow* cases that employed the merits-first-immunity-second sequence, five found that plaintiff's allegations failed to rise to a constitutional

³⁷² See, e.g., *Saucier*, 533 U.S. 193, 201 (2001).

³⁷³ *Paul v. Davis*, 424 U.S. 693, 701 (1976).

³⁷⁴ 438 U.S. 478, 517-30 (1978) (Rehnquist, J., dissenting).

³⁷⁵ *Id.* at 529-30.

violation.³⁷⁶ The sixth—*Wilson v. Layne*—found that the right in question was not clearly established at the time, and thus awarded qualified immunity at the second step.³⁷⁷

Chief Justice Rehnquist thus installed an anti-*Marburian* step into the qualified-immunity analysis. The threshold inquiry questioned whether the right that was allegedly violated existed at all. It had been in the making long before it was lifted into law in *Saucier v. Katz*. The threshold inquiry had unforeseen consequences when it was combined with another requirement, which Justice Scalia had introduced fourteen years earlier. That second requirement was that the “clearly established” right be defined at a high level of specificity.

B. *Anderson v. Creighton’s Restriction of the Remedy: “Clearly Established” as “Highly Specific”*

*Anderson v. Creighton*³⁷⁸ introduced a simple, yet contentious notion into qualified-immunity law: When *Harlow* said the right had to be “clearly established,” it meant that the right had to be *specific*. Here, a federal agent entered and searched the Creighton family’s home without a warrant, mistakenly believing that he would find a bank robber there.³⁷⁹ In the ensuing *Bivens* litigation for an unreasonable search, the Eighth Circuit held that the

³⁷⁶ See *Graham v. Connor*, 490 U.S. 386, 396 (1989) (holding that police excessive-force claims in the stop, search, or seizure context had to be exclusively brought under the Fourth Amendment, thus foreclosing plaintiff’s excessive-force claim under substantive due process); *Siegert v. Gilley*, 500 U.S. 226, 230-31 (1991) (holding that the defendant was entitled to qualified immunity because plaintiff never allege a cognizable constitutional claim); *Hunter v. Bryant*, 502 U.S. 224, 227-28 (1991) (holding that the undisputed facts established probable cause for arrest); *County of Sacramento v. Lewis*, 523 U.S. 833, 854 (1998) (holding that plaintiff’s alleged facts failed to meet the proper fault standard in a substantive due process claim in a high-speed police chase resulting in death); *Conn v. Gabbert*, 526 U.S. 286, 292 (1999) (holding that a prosecutor causing an attorney to be searched while his client is questioned by a grand jury does not implicate attorney’s Fourteenth-Amendment right to freely choose his vocation).

³⁷⁷ 526 U.S. 603 (1999) (holding that bringing media reporters in the execution of a warrant in the home violated the Fourth Amendment, but was not clearly established at the time of the alleged conduct).

³⁷⁸ 483 U.S. 635 (1987).

³⁷⁹ *Id.* at 636.

Creightons' claim survived summary judgment on both the merits and under qualified immunity.³⁸⁰ According to the appeals court, the question of whether the search was lawful depended on a genuine issue of fact, and the right to be free from warrantless searches in the home absent probable cause and exigent circumstances was clearly established.³⁸¹

But the latter holding, Justice Scalia observed in *Anderson*, did not give proper effect to *Harlow*'s new standard. "[I]f the test of 'clearly established law' were to be applied at this level of generality, it would bear no relationship to the 'objective legal reasonableness' that is the touchstone of *Harlow*."³⁸² If the *Harlow* standard merely required allegations of "violation[s] of extremely abstract rights"³⁸³ (as the Eighth Circuit had defined it), it would give no additional leeway to judges to eliminate insubstantial suits against government officers at the summary judgment stage.³⁸⁴ Thus, *Harlow* required that "[t]he contours of the right . . . be sufficiently clear that a reasonable official would understand that what he is doing violates that right."³⁸⁵ This meant that an officer could be mistaken in his judgment that probable cause existed, but still receive qualified immunity when the law was not clear enough as to make his mistake unreasonable.³⁸⁶ In the Fourth-Amendment context, the allegedly violated right for qualified immunity purposes had to be *more* specific than the right violation alleged on the merits. The Court later reiterated this more-specificity requirement for in suits brought for

³⁸⁰ *Creighton v. St. Paul*, 766 F.2d 1269, 1272-76 (1985).

³⁸¹ *Id.*

³⁸² *Anderson*, 483 U.S. at 639.

³⁸³ *Id.*

³⁸⁴ *Id.* at 639-40.

³⁸⁵ *Id.* at 640.

³⁸⁶ *Id.* at 643.

unreasonable arrests³⁸⁷ and for police executing warrants with parties not reasonably related to the warrant's objective.³⁸⁸

While Justice Scalia's reasoning in *Anderson* was plausible, *Harlow* by no means compelled it. Nothing in the opinion—or Justice Powell's papers on the case³⁸⁹—suggested that “clearly established” required the right to be more *specific* than it is on the merits. Abstract rights can be clearly established (by, for example, a long, uncontroverted body of caselaw that speak to a legal principle) just as much as a specific right can be (by a case on point). The policy goal of eliminating insubstantial lawsuits at the summary judgment stage was one desideratum among several, not the overriding goal in the way that Scalia made it out to be.³⁹⁰ And nothing in *Harlow*'s reasonableness formula suggested that it could not sometimes merge with the merits standard if it, too, was a reasonableness standard. Indeed, Justice Scalia himself admitted that he had freely derived the requirement by asserting that the Court “had never suggested that the precise contours of official immunity can and should be slavishly derived from the often arcane rules of the common law.”³⁹¹

³⁸⁷ *Hunter v. Bryant*, 502 U.S. 224, 228 (1991) (per curiam) (“Even if we assumed, arguendo, that [the officers] erred in concluding that probable cause existed to arrest Bryant, [they] would nevertheless be entitled to qualified immunity because their decision was reasonable, even if mistaken.”).

³⁸⁸ *Wilson v. Layne*, 526 U.S. 603, 613-15 (1999) (finding media ride-alongs to arrest warrant executions in the home unconstitutional, but giving qualified immunity to the arresting officers because their mistake of law was reasonable).

³⁸⁹ See *supra*, note [X] and accompanying text.

³⁹⁰ *Anderson*, 483 U.S. at 639 (“Plaintiffs would be able to convert the rule of qualified immunity that our cases plainly establish into a rule of virtually unqualified liability simply by alleging violation of extremely abstract rights.”). Justice Stevens, in dissent, pointed that the majority’s “double-counting approach [of reasonableness] reflects . . . an overriding interest in unfettered law enforcement. It ascribes a far lesser importance of the privacy interest of innocent citizens . . .” *Id.* at 664 (Stevens, J., dissenting).

³⁹¹ 483 U.S. at 645. See also *Crawford-El v. Britton*, 523 U.S. 574, 611-12 (1998) (Scalia, J., dissenting) (“[O]ur treatment of qualified immunity . . . has not purported to be faithful to the common-law immunities that existed when § 1983 was enacted That is perhaps just as well. The § 1983 that the Court created in [*Monroe v. Pape*, 365 U.S. 167 (1961)] in 1961 bears scant resemblance to what Congress enacted almost a century earlier. . . . We find ourselves engaged, therefore, in the essentially legislative activity of crafting a sensible scheme

From a *Marburian* perspective, Scalia’s interpretation of the clearly-established requirement planted the seed for a momentous shift in the doctrine. Qualified immunity had originally been a defense that could be pleaded to give the courts discretion over remedies. But tethering the standard to the substantive right itself—requiring the right under qualified immunity to be more specific—essentially transformed the standard into a restriction on the rights side of the ledger. Constricting the “contours” of the right meant nothing other than curtailing the right itself in constitutional suits under § 1983. But the anti-*Marburian* thrust of *Anderson*’s “more-specific right” requirement would not show its full implications until the Court decided *Saucier v. Katz*.

C. Forging the Saucier-Two-Step

“Justice Scalia may see what you mean, but I’m not sure I do,” Chief Justice Rehnquist laconically interjected, drawing laughs from the crowd.³⁹² Justice Scalia had just thrown Deputy Solicitor General Paul Clement a bone in spelling out how the Fourth Amendment’s reasonableness standard gave officers room for error in the excessive force cases, and how the reasonableness standard of qualified immunity created additional room for error.³⁹³ But could an officer reasonably act unreasonably? And if so, should the analysis start with whether she acted reasonably or unreasonably? Those were the tangled questions that the Justices confronted when they sat for oral argument in *Saucier v. Katz*.³⁹⁴

of qualified immunities for the statute we have intended—rather than applying the common law embodied in the statute that Congress wrote.”).

³⁹² *Oral Argument: Saucier v. Katz*, OYEZ (Mar. 20, 2001), at 10:40; <https://www.oyez.org/cases/2000/99-1977>.

³⁹³ *Id.* at 10:10-:10:38.

³⁹⁴ *See generally id.*

Respondent Elliot Katz, an elderly animal rights activist, claimed that military police officers had violently thrown him into a van while he was protesting alleged animal rights abuses during a speech by Vice President Al Gore.³⁹⁵ Katz brought a *Bivens* claim for excessive force. Confronted with two reasonableness inquiries, the Ninth Circuit concluded that the merits and immunity analyses merged into one: *Graham v. Connor*³⁹⁶ had clearly established the right to be free from excessive force in the arrest context.³⁹⁷ Objective reasonableness under the Fourth Amendment, and objective reasonableness for the purposes of qualified immunity thus reduced to the same analysis.³⁹⁸

The Justices thus had to grapple with several thorny questions: Should the constitutional and immunity analyses remain distinct in excessive force cases under *Bivens* and § 1983? If so, how could they be kept separate? In what order should courts consider them? Should a jury be instructed on both at once, or separately? And given that the videotape of Katz's arrest seemed to show that the plaintiff suffered *no* force at all, was there a real need to consider those questions in the first place?³⁹⁹

The *Saucier* Court decided that the analyses should not be merged. To keep them distinct, it held that a court *first* had to determine whether defendant's alleged conduct amounted to a constitutional violation, and *second* if the right was clearly established at the time.⁴⁰⁰ The Court also emphasized that reasonableness under qualified immunity was not the

³⁹⁵ *Saucier v. Katz*, 533 U.S. 194, 197-98 (2001).

³⁹⁶ 490 U.S. 386 (1989).

³⁹⁷ *Katz v. United States*, 194 F.3d 962, 968 (1999).

³⁹⁸ *Id.*

³⁹⁹ In the words of an exasperated Justice O'Connor at oral argument: "[W]hy are we here?" *Oral Argument: Saucier v. Katz*, OYEZ (Mar. 20, 2001), at 31:49; <https://www.oyez.org/cases/2000/99-1977>.

⁴⁰⁰ *Saucier*, 533 U.S. at 201.

same as reasonableness on the merits: “clearly established” required a higher degree of specificity than on the merits.⁴⁰¹ Qualified immunity allowed officers to make reasonable mistakes when determining which laws applied to their conduct.⁴⁰²

On its face, the *Saucier* two-step seems to be a perfectly *Marburian* approach to balancing rights and remedies. “[A] court might find it necessary to set forth principles [to aid] the law’s elaboration from case to case,” while it should also “resolv[e] immunity questions at the earliest possible stage in litigation.”⁴⁰³ Upon closer examination, however, *Saucier* was the last gasp of a *Marburian* rationale that had been dissipating from qualified-immunity doctrine for decades.

3. Dead on Arrival: *Saucier*’s Internal Tensions

It took only eight years for the Supreme Court to reverse its *Saucier* holding in *Pearson v. Callahan*.⁴⁰⁴ This was because it forged two un-*Marburian* tendencies into one framework. *Anderson*’s more-specific requirement turned out to be too demanding for plaintiffs hoping to vindicate their rights, while Chief Justice Rehnquist’s sequencing requirement introduced uncertainty as to whether the allegedly violated right even existed. Now that the two requirements had been conjoined, their interaction began to corrode qualified-immunity doctrine in unanticipated ways, stripping it of its *Marburian* wisdom. On the rights-side, the post-*Saucier* Court failed to affirm the existence of constitutional rights because it regularly splintered on the scope and existence of the right. The status of that right was even more

⁴⁰¹ *Id.* at 205.

⁴⁰² *Id.* at 206.

⁴⁰³ *Id.* at 201.

⁴⁰⁴ *Pearson v. Callahan*, 555 U.S. 223 (2009).

questionable when the Court awarded qualified immunity at step two, making the constitutional ruling seem unnecessary and mere dicta. On the remedies side, *Anderson*'s more-specific requirement would slowly move the Court towards more and more grants of qualified immunity. Between *Saucier* and *Pearson*, the Court never managed to strike the *Marburian* balance of recognizing a rights violation but denying the remedy again.⁴⁰⁵

A. Problems on the Rights Side

The Court in the first decade of the new millennium was a fractured one, and it clearly showed in the constitutional merits questions of its qualified immunity decisions.⁴⁰⁶ The Court rarely agreed on the scope—or even the existence—of a constitutional right, which introduced uncertainty on the rights side of *Marbury*'s ledger. In *Hope v. Pelzer*, an Alabama state prisoner had been twice tied to a hitching post for punitive purposes, and once was left in the burning sun for hours without water or bathroom breaks.⁴⁰⁷ Justice Stevens, writing for six members of the Court, found a plain violation of the Cruel and Unusual Punishment Clause.⁴⁰⁸ Yet, Justice Thomas, in dissent and joined by two other Justices, dissected the record of what precisely was alleged against which prison guard, and remained unconvinced that a constitutional violation had occurred.⁴⁰⁹ *Groh v. Ramirez* featured a search warrant that, under

⁴⁰⁵ *Hope v. Pelzer*, 536 U.S. 730 (2002); *Groh v. Ramirez*, 540 U.S. 551 (2004); *Brosseau v. Haugen*, 543 U.S. 194 (2004); *Scott v. Harris*, 550 U.S. 372 (2007); *Morse v. Frederick*, 551 U.S. 393 (2007).

⁴⁰⁶ See generally, MARK TUSHNET, *A COURT DIVIDED: THE REHNQUIST COURT AND THE FUTURE OF CONSTITUTIONAL LAW* (2006) (discussing the late years of the Rehnquist Court); JOAN BISKUPIC, *THE CHIEF: THE LIFE AND TURBULENT TIMES OF CHIEF JUSTICE JOHN ROBERTS* (2019) (Roberts Court).

⁴⁰⁷ 536 U.S. 730, 733-35 (2002).

⁴⁰⁸ *Id.* at 736-38.

⁴⁰⁹ *Id.* at 748-52 (Thomas, J., dissenting).

“items to be seized,” simply described the residence at which to conduct the search.⁴¹⁰ The Court, again through Justice Stevens, found the warrant facially deficient for failing the particularity requirement, as specified in the constitutional text itself.⁴¹¹ Thomas again dissented on the merits and argued that a warrant failing the particularity requirement need not always amount to an unreasonable search.⁴¹²

In *Scott v. Harris*⁴¹³ and *Brosseau v. Haugen*,⁴¹⁴ Justice Stevens now found himself in dissent against conservative majorities. In *Scott*, Justice Scalia announced the rule that, when a fleeing suspect in a high-speed chase poses a threat of serious physical injury to others, a police officer acts reasonably when bumping the motorist off the road to end the chase.⁴¹⁵ Stevens penned a lone dissent attacking both the rule and its application to the facts.⁴¹⁶ In *Brosseau*, the Court did not rule on the constitutional question but opined that, when a suspect fleeing by vehicle is armed and dangerous to bystanders, it is not unreasonable to seize them by shooting at them.⁴¹⁷ Stevens again was the lone dissenter on the merits.⁴¹⁸

The most telling case, however, was *Morse v. Frederick*, in which a high school student’s banner saying “BONG HiTS 4 JESUS” at a school-sponsored event resulted in his

⁴¹⁰ 540 U.S. 551, 554-55 (2004).

⁴¹¹ *Id.* at 557-63.

⁴¹² *Id.* at 571-78 (Thomas, J., dissenting).

⁴¹³ 550 U.S. 372 (2007).

⁴¹⁴ 543 U.S. 194 (2004) (per curiam).

⁴¹⁵ 550 U.S. 372, 381-84 (2007). Indeed, Justice Scalia announced this rule after holding that, at summary judgment, unedited videotapes refuting the non-moving party’s version of events may be considered in deciding the motion. *Id.* at 377-81.

⁴¹⁶ *Id.* at 389-95 (Stevens, J., dissenting).

⁴¹⁷ 543 U.S. 194, 199-202 (2004).

⁴¹⁸ *Id.* at 210-17 (Stevens, J., dissenting).

suspension—and four different Supreme Court opinions on free speech.⁴¹⁹ Chief Justice Roberts’ bare majority held that there was no violation of the First Amendment, as schools could take steps to restrict speech that could be reasonably interpreted to advocate illegal drug use.⁴²⁰ Justice Thomas concurred, but emphasized that he believed students should have no rights on school grounds and instead be subject to the common law principle of *in loco parentis*.⁴²¹ Justice Alito concurred to emphasize the narrowness of the Court’s holding, in that it in no way restricted political speech.⁴²² Justice Stevens, in turn, dissented on the grounds that the majority had ignored basic First Amendment principles in allowing the prohibition of a “nonsense banner” that displayed a “silly message.”⁴²³ In none of the cases between *Saucier* and *Pearson* did the Court manage to speak with a clear voice on the rights question, the way it had in *Marbury*.

But disagreements on the constitutional merits were not the only issue that beset the *Saucier* framework. Even more troublingly from a *Marburian* perspective, Justice Breyer led the charge in attacking the *principal* validity of those constitutional holdings. Repeatedly, Breyer criticized that ruling on the merits was “unwise and unnecessary.”⁴²⁴ Oftentimes, he reasoned, the decisions would be “poorly presented,”⁴²⁵ “difficult,”⁴²⁶ “unusually portentous,”⁴²⁷ and so

⁴¹⁹ 551 U.S. 393 (2007).

⁴²⁰ *Id.* at 403-10.

⁴²¹ *Id.* at 413-16 (Thomas, J., concurring).

⁴²² *Id.* at 422-25 (Alito, J., concurring).

⁴²³ *Id.* at 435, 444 (Stevens, J., dissenting).

⁴²⁴ *Id.* at 425 (Breyer, J., concurring). *See also* *Brosseau v. Haugen*, 543 U.S. 194, 201 (2004) (Breyer, J., concurring).

⁴²⁵ *Morse*, 551 U.S. at 431 (Breyer, J., concurring in part and dissenting in part).

⁴²⁶ *Brosseau*, 543 U.S. at 201 (Breyer, J., concurring).

⁴²⁷ *Morse*, 551 U.S. at 427 (Breyer, J., concurring in part and dissenting in part).

“fact dependent that the result will be confusion rather than clarity.”⁴²⁸ Forcing courts to decide those questions conflicted with the principle of constitutional avoidance, and would lead to constitutional rulings of questionable value.⁴²⁹ Institutionally, it also made little sense not to speak with one voice on the question, especially when, as in *Morse*, “decid[ing a] case on the ground of qualified immunity . . . would be *unanimous*.”⁴³⁰ Justices Stevens⁴³¹ and Ginsburg⁴³² joined the charge. Chief Justice Rehnquist’s introduction of the threshold inquiry on the existence or non-existence of a constitutional right worked its anti-*Marburian* results. It splintered the Court on constitutional rights, left the existence and scope of those rights uncertain, and induced a liberal stalwart on the Court to attack the general validity of such holdings and conclude with the clarion call: “I would end the failed *Saucier* experiment now.”⁴³³

B. Problems on the Remedies Side

When the Court reached step two in the *Saucier* framework in *Hope* (denying immunity), *Grob* (denying immunity),⁴³⁴ and *Brosseau* (granting immunity), it disagreed on the immunity question, too, albeit not as vigorously. In *Hope*, the Stevens majority corrected the

⁴²⁸ *Scott v. Harris*, 500 U.S. 372, 388 (2007) (Breyer, J., concurring).

⁴²⁹ *Morse*, 551 U.S. at 431 (Breyer, J., concurring in part and dissenting in part); *Scott* 500 U.S. at 388 (Breyer, J., concurring).

⁴³⁰ *Morse*, 551 U.S. at 431 (Breyer, J., concurring in part and dissenting in part).

⁴³¹ *Bunting v. Mellen*, 541 U.S. 1019, 1019 (2004).

⁴³² *Brosseau*, 543 U.S. at 200-02 (joining Breyer).

⁴³³ *Morse*, 551 U.S. at 432 (Breyer, J., concurring in part and dissenting in part). As will be discussed below, after the *Saucier* order of battle was abandoned, the *Anderson* requirement for a *more* particular articulation of the right than would be necessary on the merits created a ready incentive to frequently dodge the constitutional question and leave it in limbo. *Infra*, Section III.C.iii.

⁴³⁴ *Hope* and *Grob* were the last cases, as of June 2020, in which the Court denied qualified immunity and ordered that a case to go to trial.

Eleventh Circuit’s interpretation of the clearly-established standard to be more permissive than requiring “materially similar” facts in a previous case. “[O]fficials can still be on notice that their conduct violates established law even in novel factual situations.”⁴³⁵ Justice Thomas, in dissent for three Justices, argued that the majority had misconstrued the Eleventh Circuit’s standard and that his own construction of the allegedly violated right was not clearly established at the time. He also distinguished the majority’s circuit court law and other legal authorities and drew on others to contest that the law was *not* clearly established at the time.⁴³⁶ In *Grob*, four Justices, in two dissents, would have granted qualified immunity to the searching officer relying on the facially deficient warrant because his belief in its validity was reasonable.⁴³⁷ In *Brosseau*, Justice Stevens dissented from the grant of qualified immunity by arguing the question should have been decided by a jury.⁴³⁸ In *Scott* and *Morse*, however, there was no disagreement on the grant of qualified immunity.

But while members of the Court could more readily agree on the outcome of the qualified immunity question, the pressure it put on the *Saucier* framework was even stronger than the merits question. The worry arose in cases where defendant lost on constitutional grounds, but won judgment by receiving qualified immunity. Could such a defendant appeal the constitutional ruling? Allowing such reviews would go against the Court’s longstanding prudential rule that it “review[ed] judgments, not statements in opinions.”⁴³⁹ In *Bunting v.*

⁴³⁵ *Hope v. Pelzer*, 536 U.S. 730, 741 (2002).

⁴³⁶ *Id.* at 752-63. These strategies of contesting how clearly established a law was at a particular time will be discussed further in Section III.c, *infra*.

⁴³⁷ *Groh v. Ramirez*, 540 U.S. 551, 566-68 (2004) (Kennedy, J., dissenting); *id.* at 579 (Thomas, J., dissenting).

⁴³⁸ *Brosseau v. Haugen*, 543 U.S. 194, 202-03 (Stevens, J., dissenting).

⁴³⁹ *Bunting v. Mellen*, 541 U.S. 1019 (2004) (Scalia, J., dissenting from denial of certiorari) (quoting *California v. Rooney*, 483 U.S. 307, 311 (1987))

Mellen, which was a denial of certiorari, a total of five Justices across the opinion the dissent agreed that “[t]he perception of unreviewability undermines adherence to the sequencing rule we have created.”⁴⁴⁰ Justice Breyer—occasionally joined by Justices Scalia and Ginsburg—had voiced this concern in *Brosseau*,⁴⁴¹ and reiterated it in *Scott*⁴⁴² and *Morse*.⁴⁴³ The remedy prong thus put pressure on the constitutional prong when qualified immunity was awarded. Was the constitutional ruling unreviewable law? Or, since it was immaterial to the judgment, was it dicta without binding effect? The latter view was most poignantly put by Judge Pierre Leval, who called the *Saucier* rule “a puzzling misadventure in constitutional dictum.”⁴⁴⁴

C. *The Center Did Not Hold: Sliding Into Pearson*

In sum, the *Saucier* framework was inherently un-*Marburian*. Each prong interfered with the other in a way that rendered the framework toothless. On one side, questioning the existence of a right at the first step invited disagreement that would rarely lead to law that was clearly established for the purposes of step two. Requiring all lower courts to pass on difficult constitutional questions certainly did not facilitate the law’s clarification. On the other side, resolving cases at step two left it unclear what the legal status of the constitutional ruling at step one was. When the judgment turned on the qualified-immunity determination, was the constitutional ruling—often eagerly passed over by overburdened courts—advisory?

⁴⁴⁰ *Id.* at 1025 (Scalia, J., dissenting from denial of certiorari). *Accord id.* at 1019 (Stevens, J.).

⁴⁴¹ *Brosseau*, 543 U.S. at 201 (Breyer, J., concurring).

⁴⁴² *Scott v. Harris*, 550 U.S. 372, 388 (Breyer, J., concurring).

⁴⁴³ *Morse v. Frederick*, 551 U.S. 393, 432 (Breyer, J., concurring in part and dissenting in part).

⁴⁴⁴ Pierre N. Leval, *Judging Under the Constitution: Dicta About Dicta*, 81 N.Y.U. L. REV. 1249, 1275 (2006).

This structural tension undermined the integrity of the judicial process itself by potentially leaving appellate court rulings immune to review. If government defendants who had won judgment on qualified-immunity grounds could not appeal an adverse constitutional ruling, what were the implications for judicial supremacy—the very concept that *Marbury v. Madison* had created? Even if that grand menace could be managed, *Saucier*’s mandatory constitutional step treaded on *Marbury*’s wisdom of maintaining judicial discretion and flexibility, and exercising judicial power sparingly.

Meanwhile, outside pressure on the Court began to grow. As § 1983 suits against police officers for excessive force, warrantless entry, and unreasonable searches began to multiply, so did the briefing by civil rights groups, police orders, and federal and state government actors.⁴⁴⁵ Academics, too, increasingly to criticize qualified immunity for its difficulty, favoritism of government officials to the detriment of constitutional rights, and legal infirmity.⁴⁴⁶ Most importantly, federal judges themselves began to criticize and skirt the *Saucier* regime.⁴⁴⁷ But the framework’s inherent dilemmas, and the pushback that the Court felt from within and without the judiciary did not create any efforts to rethink and reform qualified immunity. Instead, it created the even more anti-*Marburian* temptation to retreat from the constitutional question altogether when it was possible, resolve cases on qualified immunity

⁴⁴⁵ See, for example, the docket sheets of *Saucier v. Katz*, *Scott v. Harris*, *Pearson v. Callahan*, and many other cases of the time. This kind of attention by interest groups outsized that of controversial cases such as *Scheuer v. Rhodes*, *Harlow v. Fitzgerald* and *Nixon v. Fitzgerald*, and *Mitchell v. Forsyth*.

⁴⁴⁶ Chaim Saiman, *Interpreting Immunity*, 7 U. PA. J. CONST. L. 1155 (2005); Mark R. Brown, *The Fall and Rise of Qualified Immunity: From Hope to Harris*, 9 NEV. L.J. 185 (2008); John C. Jeffries, Jr., *Reversing the Order of Battle in Constitutional Torts*, 2009 SUP. CT. REV. 115 (2009); John C. Jeffries, Jr., *What’s Wrong With Qualified Immunity?*, 62 FLA. L. REV. 851 (2010). See also Leval, *supra* note 443. Of course, academics had criticized qualified immunity much earlier. See, e.g., Rudovsky, *supra* note 93; Armacost, *supra* note 95. But the unworkability and questionable legal status of constitutional rulings was new.

⁴⁴⁷ *Pearson v. Callahan*, 555 U.S. 223, 234-35 (2009) (citing cases sidestepping the constitutional question).

grounds, and leave the constitutional law in limbo. In *Pearson v. Callahan*, the Court gave in to this temptation.

IV. *MARBURY TO THE WINDS: PEARSON V. CALLAHAN AND BEYOND*

1. *Pearson's Retreat*

Justice Souter was not letting counsel Peter Stirba off the hook. “I will admit, that is the most astonishing view of probable cause I have heard in this courtroom. Of course they had [probable cause].”⁴⁴⁸ In *Pearson v. Callahan*, police officers entered a Utah home as part of a drug sting, after a confidential informant had already entered it to purchase a small quantity of methamphetamines.⁴⁴⁹ Two hours after the informant made his purchase and returned to them, they made a warrantless entry.⁴⁵⁰ As other arguments for exceptions to the warrant requirement had failed, counsel put the full weight of his argument on extending the consent-once-removed doctrine to confidential informants—which the Tenth Circuit had held did not apply in its jurisdiction, creating a split with the Sixth, Seventh, and Ninth Circuits.⁴⁵¹

But Justice Souter would not let go of the obvious point: why extend the exception when probable cause was clearly established by the informant’s lawful entry and police had ample time to obtain a warrant? After several minutes of persistent questioning, Stirba had to concede that he “d[id]n’t know how to answer the question any more than [he had].”⁴⁵² He was rescued by Chief Justice Roberts who reminded him that he did not have to “prove that

⁴⁴⁸ *Oral Argument: Pearson v. Callahan*, OYEZ (Oct. 14, 2008), at 12:38; <https://www.oyez.org/cases/2008/07-751>.

⁴⁴⁹ *Pearson*, 555 at 227.

⁴⁵⁰ *Id.* at 228.

⁴⁵¹ *Oral Argument: Pearson v. Callahan*, OYEZ (Oct. 14, 2008), at 4:30-5:18; <https://www.oyez.org/cases/2008/07-751>.

⁴⁵² *Id.* at 14:26.

[he was] right” on the constitutional question, only that “the contrary principle was not clearly established.”⁴⁵³ Why engage the difficult constitutional question when, in any event, qualified immunity shielded the officers from liability, and the Court from another contentious ruling?

Pearson confronted the Court with another thorny constitutional issue, but an easy qualified immunity question. And yet the two could not be kept separate. And it faced yet another unanticipated scenario in which the constitutional and immunity analyses could not be kept separate. If the consent-once-removed doctrine applied to confidential informants, clearly established law would authorize the warrantless entry. If the doctrine did not apply, the officers would have violated clearly established law, as no exception to the warrant requirement would have authorized their entry. Resolving the constitutional question either way would decide the immunity question.⁴⁵⁴ The Court decided it was time to shed the strictures of *Saucier*. From *Pearson* onwards, “[t]he judges of the district courts and the courts of appeals [were] permitted to exercise their sound discretion in deciding which of the two prongs of the qualified immunity analysis should be addressed first.”⁴⁵⁵ A unanimous Court promptly exercised that discretion to sidestep the Fourth Amendment question and granted qualified immunity on the second prong alone.⁴⁵⁶

2. *Pearson’s Dodge*

⁴⁵³ *Id.* at 14:39.

⁴⁵⁴ The only way to keep the immunity analysis distinct was to leave the constitutional question *undecided* and accepting that the law was not clearly established.

⁴⁵⁵ *Pearson v. Callahan*, 555 U.S. 223, 236 (2009)

⁴⁵⁶ *Id.* at 243-45.

Pearson is not only interesting for the arguments it engaged in justifying its retreat from the *Saucier* two-step, but also for the arguments it chose not to engage. All of the reasons that the Court marshalled in favor of the qualified immunity's newfound flexibility were framed as interests squarely within the domain of the judiciary. *Saucier*'s rule wasted resources by requiring parties to litigate and courts to adjudicate the constitutional question.⁴⁵⁷ It risked confusing lower courts when the constitutional rule in question depended on other cases currently pending on appeal or interpretations of ambiguous state laws. It often led to premature adjudication of constitutional claims at summary judgment, when the factual record was thin and the parties' briefing was weak.⁴⁵⁸ It risked insulating constitutional rulings from appeal when a defendant lost on the constitutional issue, but won judgment on qualified immunity.⁴⁵⁹ It departed from the principle of constitutional avoidance.⁴⁶⁰ And it infantilized lower courts by prescribing a rigid sequence of steps.⁴⁶¹ *Saucier*'s goal of "support[ing] the Constitution's 'elaboration from case to case'"⁴⁶² and *Harlow*'s original concern of balancing "the need to hold public officials accountable when they exercise power irresponsibly and the need to shield officials from harassment, distraction, and liability when they perform their duties reasonably"⁴⁶³ were mentioned in passing, but not treated in depth.

⁴⁵⁷ *Id.* at 236.

⁴⁵⁸ *Id.* at 237.

⁴⁵⁹ *Id.* at 238.

⁴⁶⁰ *Id.* at 240.

⁴⁶¹ *Id.* at 242.

⁴⁶² *Id.* at 232 (quoting *Saucier v. Katz*, 533 U.S. 194, 201 (2001)).

⁴⁶³ *Id.* at 231.

Indeed, the *Pearson* Court left the implications for the development of constitutional rights and government effectiveness entirely unaddressed, despite extensive briefing on both sides by civil rights organizations, police organizations, the federal government, and the states.⁴⁶⁴ Nor did it take stock of how qualified immunity doctrine had shifted in recent years.⁴⁶⁵ From the outset, the *Pearson* opinion made every effort to frame qualified immunity as a “rule [that] is judge made and implicates an important matter involving internal Judicial Branch operations. Any change should come from this Court, not Congress.”⁴⁶⁶ The Court had drawn inwards. Its holding in *Pearson* further pursued that self-insulation by allowing the judiciary to retreat from constitutional questions. This came at a time when the Court’s qualified-immunity decisions had become a hotly contested battleground between civil rights groups and police interest groups and states on the issue of police overreach.⁴⁶⁷ As the Court attempted to dodge the political implications of its decisions, it allowed qualified immunity to deviate from the *Marburian* wisdom of previous decades—the very same wisdom that was needed to stay above the fray.

3. *Pearson’s* Aftermath

After *Pearson*, the story of qualified immunity is one of stagnation and retrenchment. With the abandonment of Chief Justice Rehnquist’s sequencing requirement, the Court seized on Justice Scalia’s requirement from *Anderson v. Creighton* that “clearly established” law meant

⁴⁶⁴ See *supra* note 444 and accompanying text.

⁴⁶⁵ Cf. *Malley v. Briggs*, 475 U.S. 335, 341 (1986) (“As the qualified immunity defense has evolved, it provides ample protection to all but the plainly incompetent or those who knowingly violate the law.”).

⁴⁶⁶ *Pearson*, 555 U.S. at 233-34.

⁴⁶⁷ See *supra* note 444 and accompanying text.

that it had to be so specific as to make the challenged conduct’s “unlawfulness . . . apparent.”⁴⁶⁸ Kit Kinports has explained in great detail how the clearly-established standard has become more and more demanding over time. She has also tracked how the Court has repeatedly stripped back sources that count for clearly established law, while allowing more and more avenues to establish that the law is *not* clearly established.⁴⁶⁹

Beyond the clearly-established standard, the doctrine itself has been refined and developed only four more times since then: *Camreta v. Greene* held that, even when a government official wins on qualified immunity-grounds but loses on the constitutional grounds, the Supreme Court may still review the appellate court’s constitutional ruling.⁴⁷⁰ *Ashcroft v. al-Kidd*, in passing, expanded that rationale to mean that, the Supreme Court has discretion to review both prongs of the qualified immunity analysis when an appellate court reaches them.⁴⁷¹ *Filarsky v. Delia* established the principle that a private individual whom the government hires temporarily to carry out its work receives the same qualified immunity protections as her permanently employed counterparts.⁴⁷² Lastly, *Plumbhoff v. Rickard* clarified that, as a general matters, denials of qualified immunity are immediately appealable, so long as they are based on a legal determination.⁴⁷³ *Mitchell v. Forsyth*, in other words, was the norm, while *Johnson v. Jones* was the exception.⁴⁷⁴

⁴⁶⁸ *Anderson v. Creighton*, 483 U.S. 635, 640 (1987).

⁴⁶⁹ See Kinports, *supra* note 76.

⁴⁷⁰ 563 U.S. 692, 700-10 (2011).

⁴⁷¹ 563 U.S. 731, 735 (2011).

⁴⁷² 566 U.S. 377, 383-91 (2012).

⁴⁷³ 572 U.S. 765, 771-79 (2014).

⁴⁷⁴ *Id.* at 771-72.

But given how little doctrinal innovation there has been since *Pearson*, the best way to tell qualified immunity's story over the last decade may be through numbers. Out of twenty-two cases, all found for the defendant official.⁴⁷⁵ Twenty of them granted qualified immunity, while one resolved the case by finding no constitutional violation,⁴⁷⁶ and another remanded on the question of whether plaintiff could bring a *Bivens* action in the first place.⁴⁷⁷ No case denied qualified immunity. Out of the twenty cases granting immunity, seventeen reversed an appellate court's denial of it,⁴⁷⁸ and three left a grant of immunity in place.⁴⁷⁹ In its twenty-two cases since *Pearson*, the Court reached the constitutional question six times. Four times, it found no violation,⁴⁸⁰ once it found one,⁴⁸¹ and once it split the difference between several constitutional claims in *Safford v. Redding*.⁴⁸² Only once, in other words, did the Court ever come close to a *Marburian* move. On another occasion, it equitably vacated the constitutional ruling because the opposing party's loss of interest in the matter had made the case moot.⁴⁸³

⁴⁷⁵ See *infra*, notes 409–411.

⁴⁷⁶ *Nieves v. Bartlett*, 139 S. Ct. 1715 (2019).

⁴⁷⁷ *Hernandez v. Mesa*, 137 S. Ct. 2003 (2017).

⁴⁷⁸ *Ashcroft v. al-Kidd*, 563 U.S. 731, 741–45 (2011); *Messerschmidt v. Millender*, 565 U.S. 535, 553–55 (2012); *Ryburn v. Huff*, 565 U.S. 469, 474–77 (2012); *Filarsky v. Delia*, 566 U.S. 377, 393–94 (2012); *Reichle v. Howards*, 566 U.S. 658, 664–69 (2012); *Stanton v. Sims*, 571 U.S. 3, 4–9 (2013); *Wood v. Moss*, 572 U.S. 744, 756–59 (2014); *Plumbhoff v. Rickard* 572 U.S. 765, 779–81 (2014); *Carroll v. Carman*, 574 U.S. 13, 17–20 (2014); *Taylor v. Barks*, 575 U.S. 822, 824–26 (2015); *City of San Francisco v. Sheehan*, 135 S. Ct. 1765, 1776–78 (2015); *Mullenix v. Luna*, 136 S. Ct. 305, 308–14 (2015); *White v. Pauly*, 137 S. Ct. 548, 550–53 (2015); *Ziglar v. Abbasi*, 137 S. Ct. 1843, 1865–69 (2017); *District of Columbia v. Wesby*, 138 S. Ct. 577, 589–93 (2018); *Kisela v. Hughes*, 138 S. Ct. 1148, 1152–55 (2018); *City of Escondido v. Emmons*, 139 S. Ct. 500, 501–04 (2019).

⁴⁷⁹ *Safford v. Redding*, 557 U.S. 364, 377–79 (2009); *Camreta v. Greene*, 563 U.S. 692, 710–13 (2011); *Lane v. Franks*, 573 U.S. 228, 243–47 (2014).

⁴⁸⁰ *al-Kidd*, 563 U.S. at 736–41; *Plumbhoff*, 572 U.S. at 773–79; *Sheehan*, 135 S. Ct. at 1774–76; *Nieves*, 139 S. Ct. at 1721–28.

⁴⁸¹ *Lane*, 573 U.S. at 238–43.

⁴⁸² 557 U.S. at 372–77.

⁴⁸³ *Camreta*, 563 U.S. at 710–13.

The picture is even starker when one looks at suits against police officers for accused of excessive force or warrantless entry. All eleven cases have reversed the appellate court's denial of qualified immunity.⁴⁸⁴ Two reached the constitutional issue, only to find no violation.⁴⁸⁵ Moreover, seven out of nine post-*Pearson* summary reversals in *per curiam* form came in this subset of cases.⁴⁸⁶ The numbers tell a story that current scholarship,⁴⁸⁷ advocacy groups,⁴⁸⁸ and media outlets⁴⁸⁹ near-unanimously confirm: the Court has abandoned its *Marburian* middle ground and come out against constitutional rights. What was initially a defense in a constitutional suit against government officials is now a “one-sided approach [that has] transform[ed] the doctrine into an absolute shield for law enforcement officers.”⁴⁹⁰

V. UPSHOTS: DECIDING TO DECIDE

This chapter closes with a few upshots on the nature of judicial power and the role of the judiciary in the American polity. These remarks must necessarily be tentative, as qualified immunity is only one case study under the *Marburian* framework.

⁴⁸⁴ *Messerschmidt v. Millender*, 565 U.S. 535, 553-55 (2012); *Ryburn v. Huff*, 565 U.S. 469, 474-77 (2012); *Stanton v. Sims*, 571 U.S. 3, 4-9 (2013); *Plumbhoff*, 572 U.S. at 779-81; *Carroll v. Carman*, 574 U.S. 13, 17-20 (2014); *Sheehan*, 135 S. Ct. at 1776-78; *Mullenix v. Luna*, 136 S. Ct. 305, 308-14 (2015); *White v. Pauly*, 137 S. Ct. 548, 550-53 (2015); *District of Columbia v. Wesby*, 138 S. Ct. 577, 589-93 (2018); *Kisela v. Hughes*, 138 S. Ct. 1148, 1152-55 (2018); *City of Escondido v. Emmons*, 139 S. Ct. 500, 501-04 (2019).

⁴⁸⁵ *Plumbhoff*, 572 U.S. at 773-79; *Sheehan*, 135 S. Ct. at 1774-76.

⁴⁸⁶ *Ryburn*, 565 U.S. at 469; *Stanton*, 571 U.S. at 3; *Carroll*, 574 U.S. at 13; *Mullenix*, 136 S. Ct. at 305; *White*, 137 S. Ct. at 548; *Kisela*, 138 S. Ct. at 1148; *Emmons*, 139 S. Ct. 500. The other two *per curiam* decisions involved border patrol agents, *Hernandez v. Mesa*, 137 S. Ct. 2003 (2017), and prison guards, *Taylor v. Barks*, 575 U.S. 822 (2015).

⁴⁸⁷ See *supra*, Introduction.

⁴⁸⁸ See *supra* note 89 and accompanying text.

⁴⁸⁹ See *supra*, Introduction.

⁴⁹⁰ *Kisela v. Hughes*, 138 S. Ct. 1148, 1162 (Sotomayor, J., dissenting).

Private individuals suing government officials in their individual capacity is a unique institutional scenario. It is unlike a criminal prosecution (in which plaintiff and defendant are reversed), a suit against the federal government (to which the government has to consent),⁴⁹¹ or a suit against a state (which are barred by the Eleventh Amendment).⁴⁹² Here, an ordinary member of the people, from whom the law's authority ultimately derives, marshals that authority against the powers entrusted with executing the law. The individual takes that controversy to court in order to vindicate *Marbury's* maxim that she lives in "a government of laws, and not of men."⁴⁹³ Individual suits against government officers thus touch on the fundamental question of where the power to give effect to the law ultimately resides. Through individual litigants, the people as such appear as an institutional actor that asks: Is the government, too, bound by the law?

The history of qualified immunity as a *Marburian* move gone wrong shows that, initially, the judiciary seriously engaged that question. *Pierson v. Ray* declined strict liability for police officers in favor of a fault standard;⁴⁹⁴ *Scheuer v. Rhodes* declined to apply Eleventh-Amendment sovereign immunity and ordered the matter to proceed to trial;⁴⁹⁵ *Butz v. Economou* declined absolute immunity for federal executive officials.⁴⁹⁶ Together with the steady objectification of the standard, the pre-*Harlow* Court showed an express interest in adjudicating individuals' suits against government officials into court. By taking a *Marburian* middle path

⁴⁹¹ *United States v. Shaw*, 309 U.S. 495, 500-01 (1940). See also Note, *Government Tort Liability*, 111 HARV. L. REV. 2009, 2009 (1998).

⁴⁹² *In re New York*, 256 U.S. 490, 497 (1921).

⁴⁹³ *Marbury v. Madison*, 5 U.S. (1 Cranch) 137, 163 (1803).

⁴⁹⁴ 386 U.S. 547, 555 (1967).

⁴⁹⁵ 416 U.S. 232, 237 (1974).

⁴⁹⁶ 438 U.S. 478, 504-08 (1978)

between absolute immunity and strict liability, the Court declined to make the law a bludgeon for either party, and instead decided that the judiciary decide those cases. Indeed, the *Marburian* move of creating new judicial authority in a standoff between the people and the government boils down to this unrecognized aspect of judicial power: deciding to decide cases.⁴⁹⁷

In deciding to decide cases brought by individuals against government officials, the Court recognized the people as a legitimate institutional player, while enhancing the power of the judiciary by “rationing” the remedy.⁴⁹⁸ The Court’s early understanding of qualified immunity coheres with several institutional of the Court. It expands Alexander Bickel’s theory of the Supreme Court as a “statesman” mediating between the governmental branches, as the Court also mediates between the government and members of the populace.⁴⁹⁹ It squares with John Hart Ely’s process theory of empowering minority participation in the political process, if understood as direct vindication of individual rights.⁵⁰⁰ The Court’s early understanding of qualified immunity also coheres with Charles Black’s theory that the Court is a legitimator of the political process:⁵⁰¹ funneling a standoff between the people and government officials into the judicial forum defuses tensions, as long as the process is evenhanded.

But as qualified immunity jurisprudence progressed in its post-*Harlow* stage, the Court effectively decided to no *longer* decide those cases by rationing the remedy out of existence.⁵⁰²

⁴⁹⁷ This chapter does, of course, not dispute that courts also decide cases, pure and simple. Rather, it draws attention to the second-order decision of whether or not to decide cases as well.

⁴⁹⁸ Huq, *supra* note 112, at 2-29.

⁴⁹⁹ ALEXANDER M. BICKEL, *THE LEAST DANGEROUS BRANCH: THE SUPREME COURT AT THE BAR OF POLITICS* (1962).

⁵⁰⁰ JOHN HART ELY, *DEMOCRACY AND DISTRUST: A THEORY OF JUDICIAL REVIEW* 73-184 (1980).

⁵⁰¹ C.L. BLACK, JR., *THE PEOPLE AND THE COURT* 34 (1960).

⁵⁰² This is where my account complicates and expands on Huq’s theory: rationing remedies does not enhance judicial legitimacy in a linear fashion, but runs into a limit and undermines its legitimacy at some point.

Departing from *Marbury*'s wisdom in that way not only cuts against the above institutional theories by ceasing to mediate between the people and the government, denying minority participation in the judicial process, and refusing to do the work of legitimation. But it also subverts the judiciary's institutional interests in independence and legitimacy. One thing on which academics, public opinion, cross-ideological interest groups, and even some Justices agree is that qualified immunity is a political and institutional liability to the Court. As this chapter and many articles on the subject have shown, this loss of legitimacy has coincided with the Court's removal of both rights and remedies from reach.

The Marburian analysis of qualified immunity's history in its institutional context thus sides with Justice White. Contrary to then-Justice Rehnquist's pronouncements in his dissent in *Butz v. Economou*, it is imperative that qualified immunity doctrine "hold a middle ground."⁵⁰³ The seed for doctrinal reform may reside in *Nieves v. Bartlett*.

⁵⁰³ Butz v. Economou, 438 U.S. 478, 526 (1978) (Rehnquist, J., dissenting).

ProQuest Number: 28320491

INFORMATION TO ALL USERS

The quality and completeness of this reproduction is dependent on the quality and completeness of the copy made available to ProQuest.



Distributed by ProQuest LLC (2021).

Copyright of the Dissertation is held by the Author unless otherwise noted.

This work may be used in accordance with the terms of the Creative Commons license or other rights statement, as indicated in the copyright statement or in the metadata associated with this work. Unless otherwise specified in the copyright statement or the metadata, all rights are reserved by the copyright holder.

This work is protected against unauthorized copying under Title 17,
United States Code and other applicable copyright laws.

Microform Edition where available © ProQuest LLC. No reproduction or digitization of the Microform Edition is authorized without permission of ProQuest LLC.

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 - 1346 USA